

Health, knowledge and the demand for medical care : an econometric analysis

Citation for published version (APA):

Doorslaer, E. K. A. (1987). *Health, knowledge and the demand for medical care : an econometric analysis*. [Doctoral Thesis, Maastricht University]. Rijksuniversiteit Limburg.
<https://doi.org/10.26481/dis.19871217ed>

Document status and date:

Published: 01/01/1987

DOI:

[10.26481/dis.19871217ed](https://doi.org/10.26481/dis.19871217ed)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

health, knowledge and the demand for medical care

HEALTH, KNOWLEDGE AND THE DEMAND FOR MEDICAL CARE

AN ECONOMETRIC ANALYSIS

Proefschrift

ter verkrijging van de graad van doctor in de Sociale Wetenschappen
aan de Rijksuniversiteit Limburg te Maastricht, op gezag van de
Rector Magnificus, Prof.Dr. F.I.M. Bonke,
volgens het besluit van het College van Dekanen,
in het openbaar te verdedigen op donderdag 17 december 1987
om 16.00 uur .

door

Edy Karel Alina Van Doorslaer

geboren te Willebroek op 24 februari 1958

Promotor: Prof.Dr. F.F.H. Rutten
Referenten: Prof.Dr.Ir. A. Kapteyn
Prof.Dr. F.Ch. Palm
Prof.Dr. W.P.M.M. van de Ven

CONTENTS

Acknowledgements

Chapter 1.	Introduction	1
1.1.	The subject of analysis	1
1.2.	The relevance of the subject matter	2
Chapter 2.	Economic models of health and medical care decision-making	5
2.1.	Introduction	5
2.2.	Static models of the demand for medical care	6
2.3.	Static models of the demand for health	7
2.4.	Lifecycle models of the demand for health	9
2.5.	Stochastic models of the demand for health	16
2.5.1.	Uncertainty with respect to health	16
2.5.2.	Uncertainty with respect to the effectiveness of medical care	19
2.6.	Conclusions and implications for research strategy	22
	Notes	25

Part I. CONSUMER INFORMATION AND THE DEMAND FOR MEDICAL CARE

Chapter 3.	The demand for information on the quality of medical care	29
3.1.	Introduction	29
3.2.	Asymmetry of information in health care	29
3.3.	A Bayesian approach to the derived demand for therapeutic information	31
3.3.1.	The Kihlstrom model of Bayesian consumer behaviour	31
3.3.2.	Comparative static predictions of the demand for information	35
3.4.	From theoretical to empirical model	38
3.4.1.	Joint product problems	38
3.4.2.	Income, prices, expectations and preferences	39
3.4.3.	Effects of information manipulation	40
3.5.	Conclusions and assessment of the theoretical model	44
	Notes	45
Chapter 4.	Empirical analysis	46
4.1.	Description of data and Belgian health care characteristics	46

4.2.	Measurement of health	47
4.2.1.	Methodology: the LISREL model	47
4.2.2.	A latent health vector	51
4.2.2.1.	Model specification and testing	51
4.2.2.2.	Model modification	56
4.2.3.	A MIMIC health model	58
4.3.	Measurement of information	61
4.3.1.	Previous studies	61
4.3.2.	A MIMIC model of consumer information	62
4.3.2.1.	Medical knowledgeability	63
4.3.2.2.	Perceived quality of GP's	66
4.4.	A structural equation model of the demand for information and medical care	70
4.4.1.	Specification of the demand equations	71
4.4.2.	Estimation results	74
4.4.3.	Robustness analysis	80
4.4.3.1.	Sample selection	81
4.4.3.2.	Model structure	83
4.5.	Conclusions and discussion	85
	Notes	89

Part II. EDUCATION AND HEALTH 91

Chapter 5.	Economic analysis of the schooling-health relationship	93
5.1.	Introduction	93
5.2.	The effect of education on health demand and production	95
5.3.	Methodological issues: a survey of previous research	99
5.3.1.	Reverse causality (B)	100
5.3.2.	Third factor (C)	101
5.3.3.	Intervening variables (D)	103
5.4.	Conclusions	105
	Notes	107
Chapter 6.	The marginal health productivity of schooling: a longitudinal latent variable model	108
6.1.	Introduction	108
6.2.	Data	108
6.3.	Measuring health changes over time	110
6.3.1.	Variability of health indicators	110
6.3.2.	Stability of latent health status	112
6.4.	A structural equation model of health production	117
6.4.1.	The schooling-health association	117
6.4.2.	Cumulative and dynamic effects	118
6.4.3.	Direct and indirect effects	122
6.5.	Conclusions	125
	Notes	127

Chapter 7.	Use-related health depreciation and education: a fixed-effects model	128
7.1.	Data	128
7.2.	Measurement of health	129
7.3.	Analysis of covariance	133
	7.3.1. Dynamic effects	133
	7.3.2. Cumulative effects	140
7.4.	Interaction effects	142
7.5.	Conclusions	146
	Notes	149
Chapter 8.	General summary and conclusions	150
References		157
Appendices		164

Acknowledgements

Although there is only one author on the front page of this book, he could never have written it without the help and encouragement of a large number of people. In chronological order, I would like to thank all those who have contributed to its accomplishment in some way or another.

Without Walter Nonneman I would probably never have heard of anything like 'health economics', let alone become interested in it. It was his conviction that a good Ph.D. in economics has to start with a study abroad that provided the stimulus for me to try to go to the University of York. The financial support of the CIM-fellowship made it possible to realize these plans. During my stay there, graduate supervisors Alan Williams and Alan Maynard made sure that this interest was further sharpened and that I returned even more enthusiastic about the subject than when I left. Although I can say that the roots of this book lie in Antwerp and York, most of the work was done in Maastricht at the Department of Health Economics under the supervision of Frans Rutten. I am very grateful to him for encouragement and advice in all stages of this research, and for giving me the opportunity to work in a small but stimulating team of researchers in the project on Determinants of Health Care Utilization. I owe many thanks to José Geurts, for intelligent assistance (and shared frustration) with the data processing and analysis; to Richard Janssen, for the transmission of some of his energy when it was needed; to Ellen Breevoort and Jan van Emmerik for skillful research assistance in a ready-while-you-wait-style; and to Frans Tan and Arnold Kester for their numerous valuable advices on statistical problems.

Furthermore, I would like to thank the referees for their comments and suggestions for improvement on various parts of the book; Ger Haan for this thorough proofreading of the manuscript and common-sense advice; Jaap van den Berg (CBS) for his willingness to participate in the research of Chapter 7 and, last but not least, Brigitte Kerbusch. It would be a severe understatement to say that she typed the text because she really processed it fast and accurately, thereby transforming a sometimes dreadful manuscript into a perfect-looking document.

Thanks are also due to the Belgian National Research Program for the Social Sciences, the Nederland Oké Project and the Netherlands Central Bureau of Statistics for letting me use their data files and to the Intercollegiate for Doctoral Studies in Management Science for financial support.

Finally, I thank my parents for investing in my health and education, Gerd for her support and encouragement throughout and my sons Stijn and Joris for letting me sleep during so many nights when I badly needed it.

Chapter 1. Introduction

"The greatest potential for improving health lies in what we do and don't do to and for ourselves" (Fuchs, 1974)

1.1. The subject of analysis

This thesis attempts to establish whether, how and to what extent an individual's knowledge affects his or her decisions with respect to health behaviour in general and the use of medical care in particular. The approach is economic in the sense that the (neo-classical) paradigm of rational behaviour under constraints is used as the basic underlying theory. Within the field of economics, the subject is located on the boundary between the economics of health and the economics of education, sometimes placed under the common label of the "economics of human resources" (Blaug, 1970, p. 317). The theoretical part also draws on some developments in the theory of consumer behaviour under uncertainty.

Although the analysis encompasses the three relationships between essentially three main variables of interest (health, knowledge and medical consumption depicted in Figure 1.1) we have concentrated on two of them in particular.

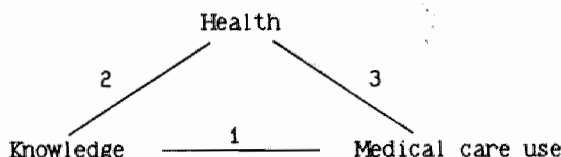


Figure 1.1. Basic relations under study

In Chapter 2, where we review the economic models of the demand for health and medical care, we also make more explicit the reasons for

subdividing the remainder of this study into two main parts.

The first part then concentrates on relationship 1 between consumer information and the demand for medical care, while controlling for health status. Information asymmetry between the patient/demander and the physician/supplier is formally modelled using a Bayesian approach. Important questions we seek to answer include: How does the consumer's medical knowledge affect his demand for medical care? Does a patient's prior information affect his response to medical advice and how does this responsiveness vary with respect to his perception of such advice?

The second part of this study is concerned with relationship 2 between individual health and education level (as a proxy for knowledge and skills). The focus there is on the possible causality running from more schooling to better health. In the human capital framework, decisions with respect to health and education are considered as investment behaviour in a lifetime perspective. Education is assumed to enhance the individual's ability to improve his health by investing time and resources in the use of medical care (the technical efficiency effect) or to actively prevent health from deteriorating by adopting a healthy lifestyle (the allocative effect). Crucial questions addressed in this chapter are: To what extent can health inequalities be attributed to differences in schooling levels? Does the schooling-health correlation also represent causality or can it entirely be explained by an omitted third factor?

Different datasets are used in the empirical analysis to study the different sub-themes. In Part One some predictions of the theory on the demand for information are tested by means of a sample of individuals taken from the Primary Health Care Survey conducted by the Belgian National Research Program for the Social Sciences in 1976. The advantage offered by this questionnaire was that it extensively asked for respondents' health status, medical consumption, knowledge and perceptions.

For the analyses of education and health which are reported in the second part, two Dutch datasets were employed: the Nederland OKé Panel Survey carried out by the University of Limburg in 1981 and 1983, and the 1984 Health Interview Survey of the Central Bureau of Statistics. These datasources were selected because they contained a longitudinal measurement of health status (at two points in time).

1.2. Relevance of the subject matter

Opinion polls have repeatedly shown that a good health is one of the most highly rated values for a large number of people. There are also indications that the proportion of people considering good health as

the most important thing in their life has increased over the last few decades.¹ This is not so surprising because according to the definition of the World Health Organization ("a state of complete physical, mental and social well-being") health comes very close to the more general concept of consumer welfare used in economics. But if health is apparently deemed so valuable, why then is there so much concern about the rising, mainly public but also private, expenditures for health care? In answering this question van de Ven (1987) argues that "it is the waste and inefficiency and an insufficient meeting of the consumer's preferences rather than the absolute level of cost which is the fundamental issue in the health care cost problem" (p. 254). He advocates a change in the incentive structure to achieve a more efficient allocation of resources within the health care sector. However, the concern does not only apply to the allocation of resources within the health care sector but also to the allocation between health care and other determinants of health. Especially the strong and consistent positive correlation between education and overall good health has often attracted the attention: "The same research by health economists that reveals the small marginal contributions of medical care and of income to health reports a strong positive relation between health and years of schooling." (Fuchs, 1986, p. 277) This book is an attempt to find out whether more knowledgeable and more informed individuals behave differently both as consumers of medical care and as non-consumers taking decisions which may directly or indirectly affect their health status. Such an investigation seems to be timely and warranted both from a theoretical and a policy point of view. The theory of the demand for health and medical care is complemented and extended with elements borrowed from other subdisciplines within economics and subjected to new empirical tests making use of more advanced econometric methods. But the analysis also has some relevance for public policy makers who consider health education programs as one of the major instruments to achieve health targets for the year 2000 (Nota 2000, 1986). Such policies are entirely based on the assumption that more health information leads to healthier behaviour, or to put it in economic terms, that it either reduces the costs or increases the (perceived) benefits of health investment. We do not intend an evaluation of specific interventions but will rather attempt to establish at a more aggregate level whether specific medical knowledge and a general cognitive development through schooling do matter in terms of health.

Notes

1. Van de Ven (1987) reports the figures of the Netherlands Social and Cultural Planning Bureau (1984) indicating that this percentage has risen from 35 to 52% over the last 20 years. Also Adriaanse et al. (1981) find very high valuations of personal health in comparison to other items like peace, freedom, friendship, family, etc.

Chapter 2. Economic models of health and medical care decision-making

"To try to formulate models of 'demand' for medical care which exclude the direct influence of the physician, is to try to stage 'Hamlet' without the Prince of Denmark" (Evans, 1984, p. 91)

2.1. Introduction

Undoubtedly, the most distinguishing feature of the economic approach towards human behaviour of individuals from other social sciences, is the adoption of a formal framework of rational decision-making under certain constraints. Human behaviour with respect to health and utilization of medical care constitutes no exception to this rule. Ever since economists started investigating the use of health care, the micro-economic neo-classical paradigm of constrained utility maximization has been a guideline for any theoretical underpinnings of empirical research efforts. As in any other subfield of economics, there has been accelerating growth in the literature on the demand for health care, extending the basic textbook model in numerous ways to allow for various special characteristics of the health care good. Michael Grossman's (1972a) book, however, stands out as the seminal work in the theory of the demand for health and medical care. His treatment of health as a human capital stock which yields consumption and investment benefits, and medical care demand as being derived from the more fundamental demand for good health, provided the foundation for most of the work which has been published afterwards. His model will occupy a fairly central place in the survey presented in this chapter because it can be seen as a benchmark with which other models can be compared.

In what follows, we will not give a detailed description of all models which are being discussed. For the purposes of our study, we will only

emphasize the mechanisms through which education or knowledge enters each of these behavioral frameworks explicitly or implicitly. The survey is not in strict chronological order but rather structured according to three important distinctive features of the models. First, we distinguish between (mostly early) models in which medical care itself directly enters the individual's utility function and those in which the demand for medical care is **derived** from the demand for health. Secondly, we divide between **static** one-period models and those which are inherently **dynamic** because of a multi-period objective function or constraints in which changes over time are explicitly taken into account. Thirdly, the models can be classified as either **deterministic** or **stochastic** depending on whether or not uncertainty is introduced through the presence of random variables. In fact, the stochastic models can be further subdivided according to which variable is assumed to be random: health status or the marginal health product of medical care. The latter category of models makes it possible to introduce the special role of the physician in health care decision-making under conditions of **asymmetric** information. In the last section, some conclusions are drawn with respect to the choice of a theoretical model which suits best the purposes of our own study.

2.2. Static models of the demand for medical care

Some of the earlier studies were primarily intended as explorations of the price and income elasticities of the demand for medical care (for a survey, see e.g. Feldstein (1974) or Newhouse (1981)). The estimated demand equations were, explicitly or implicitly, derived from the standard utility maximization model in which purchases of medical care are traded off against those of a composite good under a budget restriction (i.e. $\max U(m, c)$ subject to $p_m M + p_c C = Y$ with M medical care, C a composite good, p_m and p_c their respective market prices and Y income). Apart from prices and income, other background variables included in the empirical equations like age, sex, health status indicators or education are assumed to be 'taste shifters'. Very often these variables proved to be important predictors of utilization but the interpretation of their effects was mostly based on ad hoc explanations. Education effects, however, were inconsistent across studies which led Newhouse (1981), in his review of these demand studies, to the remark that 'the effect of education has been more difficult to isolate (p. 92)'.

Some authors adopted this simple framework in order to concentrate on the effects of **time prices** in the demand for health care. Acton (1975) observed that health care consumption often involves significant inputs of the consumer's time which may be much larger than the monetary price of medical treatment. He therefore defines the total cost per unit of health care as the monetary cost (net of insurance) plus the required time input valued at its opportunity cost and

adjusts the budget constraint accordingly. His empirical work supports the model prediction that time prices (proxied by travel time) become more important when money user prices are low. Phelps and Newhouse (1974) showed that the elasticity of demand with respect to each of the components of the total price is approximately equal to the elasticity of demand with respect to the total cost of care per unit multiplied by the proportion of the total cost made up by each of the components.

Although some of these studies have offered useful empirical evidence on the relative effects of traditional economic variables like income and prices on the demand for medical care, their treatment of health care as a direct argument in the individual's utility function ignores the fact that the basic trade-off is between health and other fundamental objects of choice (commodities). The endogenous nature of health in decision-making was first recognized by Mushkin (1962) who regarded health as a human capital stock yielding both investment and 'pure consumption' benefits. Although this article can be seen as the first attempt to view health as an economic good, it does not present a formal model of health behaviour. This was developed first by Grossman (1972a) and will be discussed in Section 2.4. In the next section, and in the logical order of increasing complexity, we will first describe an example of a static one-period model of the demand for health.

2.3. Static models of the demand for health

In order to explore the trade-off between income and health which is often made in decisions with respect to occupational choice, Kemna (1985) developed a static one-period model of the demand for health by labour force participants. Workers are assumed to maximize utility derived from consuming 'commodities' which are produced in the household by combining goods bought in the market with own time investment (Becker, 1965, Lancaster, 1966). Medical care does not enter the utility function directly but appears in the household production function for health as one of the market inputs. Kemna (1985) shows that, under certain conditions, maximizing $U(H, C)$ is equivalent to maximizing $U(H, y)$ where y is 'full income' in the Becker (1965) sense, defined as the market value of the individual's healthy time H and C is the composite good. This objective function is maximized subject to the following constraints:

$$y = wH + R - p_m M \quad (\text{budget constraint}) \quad (2.1)$$

where w is the wage rate, R denotes other non-wage income and p_m is the market price of medical care.

$$H = H(H_0, M, S, z) \quad (\text{health production function}) \quad (2.2)$$

where H_0 is the individual's predetermined health at the beginning of

the period, M is medical care input, S is the level of schooling assumed to affect the efficiency of production and z denotes exposure to occupational health hazards as a consequence of an unhealthy work environment. The latter variable is introduced because the objective of the study was to investigate an occupational linkage in the relationship between schooling and health. It is therefore also introduced in the third constraint which describes the market clearing wage function:

$$w = w(z, H, S, A) \quad (2.3)$$

Apart from the individual's own health H and the health risks of the job (z), the wage rate is assumed to be dependent on the worker's human capital as measured by schooling (S) and experience proxied by age (A).

By solving this problem with respect to the decision variables z and M , Kemna (1985) is able to derive their demand functions. However, because health is endogenous in this model, a "demand for health" function can also be derived from the first order optimality conditions for z and M . In general and implicit form it can be specified as:

$$H = H(p_H, W^*) \quad (2.4)$$

where p_H is the shadow price of health and W^* is called "real full wealth".¹ The "optimal" health level is then not exogenously given to the individual but can be increased or decreased by changing z or M . The shadow price of health associated with z defined by $(p_H(z) = -Y_z/H_z)$ is the amount of income foregone to obtain a marginal unit increase in health by choosing a lower z . Alternatively, health can be marginally increased by foregoing the amount of income associated with a marginal increase in medical care M the shadow price equals $(p_H(M) = -Y_M/H_M)$. The shadow prices of health associated with z and with M , $p_H(z)$ and $p_H(M)$, are equal at the optimum.

Because both the shadow price (or marginal cost) of health p_H and real full wealth W^* are a function of the exogenous variables, the effects of changes in any of these exogenous variables on the demand for health can be calculated. Even though the comparative static results do not always yield unambiguous sign predictions², it is illustrative to show how the effects of schooling on health, for instance, can be decomposed into direct and indirect effects as follows (see Kemna, 1985, p. 42):

$$dH/dS = H_z(z_S) + H_M(M_S) + H_S \quad (2.5)$$

where f_i denotes the partial derivative of the variable f with respect to variable i .

The first term on the right hand side of the equation represents the 'occupation effect' of schooling on health through the demand for job hazards. The second term is indicated as the 'medical care

consumption' effect of schooling on health. Finally, the third term gives the direct 'efficient production' effect of schooling on health. Kemna's (1985) empirical work brings out that the total effect of schooling on health is positive, the largest part of which is accounted for by the direct effect H_3 . Schooling is found to have a strong negative impact on the empirical construct measuring z , but only a weak positive effect (in magnitude and significance) on the demand for doctor visits. Moreover, H_M turned out to be negative (!) and strongly significant. We will not discuss the estimation problems that were encountered in the empirical testing of this model here but refer to Chapter 5 of this thesis for that purpose.

In this section we only want to assess the merits of the household production approach to individual behaviour with respect to health. The model described above assumes that people choose their optimal level of health with reference to other desired commodities. Health is not valued above anything else. This implies that the desired health level will be adjusted downward if the marginal costs of maintaining it exceed the marginal benefits in terms of utility. In contrast to health, the level of schooling obtained is exogenously determined in this model. It is, however, no longer assumed to be a mere taste shifting factor but entered explicitly in the household production function as an efficiency-increasing factor according to standard practice in household production theory (see e.g. Michael, 1972). It is not a direct factor of production because its use in the production of health does not diminish its availability for use in other productive processes.

Although this model of the demand for health certainly has some clear advantages over the models of the demand for medical care discussed in the previous section, because it offers a better insight in the trade-offs involved in deciding whether or not to consume medical care, it is somewhat unfortunate that it does not provide an unambiguous prediction of the effect of schooling on the demand for health care. As a consequence, the determination of the 'medical care consumption effect of schooling', as referred to by the author, remains an empirical matter. Moreover, a one-period time horizon overlooks the fact that decisions about activities which may be harmful or beneficial to one's health, like choosing a job, carry over to later periods of life. Enlarging the time horizon extends the range of trade-offs that can be considered from current to future costs and benefits. This aspect of health behaviour leads us to the class of models discussed in the next section.

2.4. Life-cycle models of the demand for health

As was mentioned already in the introduction to this chapter, the most important breakthrough in the demand for health theory, has been

Grossman's (1972a, 1972b) development of a formal framework to analyze health behaviour.³ The point of departure in Grossman's work is that health can be viewed as a durable capital stock which depreciates with age but can be augmented through health investments. The fundamental difference between health capital and other human capital is that the first determines the total time available to the individual for market and non-market household production while the stock of knowledge affects household productivity. The stock of health therefore has to be distinguished from the services it generates in the form of healthy time over a certain period. The initial endowed stock of health is assumed - after early adulthood - to depreciate over time at an increasing rate, but gross investments can be produced by combining own time inputs of the consumer and market goods bought such as medical care, food, exercise, etc. As a consequence of the introduction of the stock concept, consumer decisions about health can no longer only depend on marginal costs and benefits in each period separately because benefits of health also accrue in the future. The decision framework then becomes intertemporal, extending to the rest of the life cycle. Even the length of life becomes endogenous because death occurs when the stock of health falls below a critical level: at some age the individual is no longer able or willing to compensate increased depreciation by gross health investment.

To gain some more insight into the basic model set up, it is useful to reproduce here the basic equations (see Grossman, 1972b or Muurinen, 1982): Each individual is assumed to maximize life-time utility U which is a function of C_t , a composite commodity, and h_t , the healthy time in period t :

$$U = U(C_0, \dots, C_t; h_0, \dots, h_t) \quad (2.6)$$

The relation between health services h_t and health stock H_t is given by ϕ_t , the service flow per unit stock:

$$h_t = \phi_t(H_t) \quad \text{with } \phi'_t > 0 \quad (2.7)$$

By definition, in each period net investment in the stock of health equals gross investment minus depreciation:

$$H_{t+1} - H_t = I_t - \delta_t H_t \quad (2.8)$$

where I_t denotes gross investment and δ_t the rate of depreciation during period t . The length of life T is determined by the condition that

$$H_t \geq H_{\min} \quad (2.9)$$

where H_{\min} represents a minimal stock of health which is necessary to survive. Death takes place at T when $H_t < H_{\min}$. Both I_t and Z_t are produced using time and market goods as input in the following household production functions:

$$Z_t = Z_t(x_t, T_t^Z; E) \quad (2.10)$$

'physic' rate of return) and the second term (b) is the marginal monetary rate of return of health (also called the marginal efficiency of health capital).

For further analysis, Grossman distinguishes between the 'pure investment' submodel which assumes that healthy time does not yield direct utility (i.e. term (a) is zero) and the 'pure consumption' submodel in which the monetary benefits are assumed to be negligible (i.e. term (b) equals zero). Comparative static predictions for the effects of changes in the exogenous variables (i.e. age, education, wage rate and income) on the demand for health and medical care are derived for the 'pure investment' submodel in Grossman (1972b).

We will restrict our discussion here to a verbal description of the effects of education. A higher education level enables the same increase in health to be produced with fewer inputs of both time and medical care. This implies lower marginal costs of health investment, and therefore a higher demand for health. On the other hand, the education effect on the demand for medical care is negative if the elasticity of the marginal efficiency of health capital in producing healthy time is less than unity. This means that more knowledgeable consumers are expected, *ceteris paribus*, to demand more health and less medical care. Most of Grossman's (1972a) empirical analysis is restricted to white males in the labour force with positive sick time. A number of additional assumptions had to be made regarding the explicit functional forms of the utility, investment and production functions in order to be able to derive estimable demand functions for health and a reduced form demand for medical care function. The findings support the positive health effect of education but education does not show a statistically significant effect on the demand for medical care.

One of the criticisms directed at Grossman's basic model by Muirinen (1982a) was concerned with the role of education. According to conventional practice in the household production approach, education is assumed to operate as a purely technical effect at the level of the specific production relationships. However, several authors (e.g. Welch (1970) and Michael (1972)) have argued that education also enhances allocative ability, i.e. it does not only raise the marginal product per input but may also increase knowledge about innovative production methods and therefore alter the choice of input mix: 'In the context of health, the allocative benefits of education may be seen as consisting of life-style selection, diet and exercise decisions, and the more technical benefits as showing up in more efficient production of health from better medical knowledge, more information about alternative sources of care, etc.' (Muirinen, 1982a, p. 9). To incorporate these allocative benefits of education, she introduces the concept of use-related depreciation of health by analogy with the firm's investment theory. The rate of health depreciation δ then becomes a function, not only of age, but also of the intensity with which this stock is used.⁴ The use intensity itself

is dependent on a number of environmental variables (such as stress or pollution) which also includes education. The change of the health stock over time in Muurinen's (1982a) continuous time version of Grossman's model is defined by the following equation:

$$H(t) = \frac{dH}{dt} = f(t) \cdot M(t) - \delta[t, X] H(t) \quad (2.16)$$

where $f(t) \cdot M(t)$ is the new health produced by the use of medical care $M(t)$ multiplied by its marginal health product $f(t)$. The depreciation rate δ is a function of age t and environmental variables X , including education. An advantage of the use-related depreciation device is that education (and environmental variables in general) can then be operative on health also in periods without health investment (i.e. when $M(t)$ is zero). Note also that the traditional household production framework and the role of time inputs and wages as time prices have been dispensed with in this model. Health investment has a narrow interpretation of medical care use. Despite the claim of superiority of this model over Grossman's model, the comparative static results with respect to education are practically the same, though somewhat less dependent on additional assumptions (see Muurinen, 1982a, pp 18-19). The effects of education were not examined in detail in Muurinen's (1982b) empirical analysis of Finnish Health Survey Data because of data limitations. The distinction between the hypotheses about education as an efficiency-improving or a use-related depreciation factor therefore seems to be more relevant on the conceptual than on the empirical level.

The most recent attempt to test empirically some of the predictions generated by Grossman-type models of the demand for health was made by Wagstaff (1986). He advocates the estimation of structural rather than reduced form equations of the demand for health and medical care and shows that the latter may fail to detect inconsistencies between the model predictions and the data. His theoretical model is basically a continuous time version of Grossman's original model but with a broad view of health investment as including all types of health-promoting behaviour and with use-related depreciation as affected by health-damaging behaviour and environmental influences (see also Cropper, 1981). The standard household production framework is adopted with wage rates acting as time prices and education as a productivity-increasing factor. In contrast to earlier studies which estimated either the 'pure investment' or the 'pure consumption' model, he derives empirically estimable structural and reduced form demand functions for both submodels. His findings also have important implications for the effects of education. Three of them need to be mentioned here. First of all, it is shown that the coefficient of education in the reduced form demand for medical care equation can have the right (predicted) sign for the wrong reason. If Grossman's assumption is followed that health investment I is produced by combining time and medical care according to a Cobb-Douglas, constant-returns-to-scale production function, the derived marginal cost of health investment $\pi^H(t)$ becomes a loglinear function of education (see

Wagstaff, 1986, p. 199 or Grossman, 1972a, p. 97):

$$\ln \pi^H(t) = (1 - \alpha) \ln W(t) + \alpha \cdot \ln P^m(t) + \beta E \quad (2.17)$$

where P^m is the price of medical care, α is the Cobb-Douglas parameter which indicates the share of medical care in the total cost of gross investment ($0 < \alpha < 1$), $W(t)$ is the wage rate and E is education. The parameter β should be negative if education increases health productivity. Now, in the reduced form demand for medical care equation of the pure investment model the coefficient of education becomes $-\beta(\epsilon - 1)$, where ϵ is Grossman's elasticity of the marginal

efficiency of capital. It is defined as $\epsilon = \frac{1}{1+c}$ and c is derived from the healthy days production function with decreasing marginal productivity $h(t) = b \cdot H(t)^{-c}$, with $h(t)$ healthy time, $H(t)$ is health stock and b and c are positive constants. Note that $0 < \epsilon < 1$ if $c > 0$. The education-efficiency hypothesis ($\beta < 0$) cannot simply be tested by checking the sign of the reduced form parameter because $-\beta(\epsilon - 1)$ is negative:

(a) if $\beta < 0$ and $0 < \epsilon < 1$

This would be consistent with both the education-efficiency hypothesis and diminishing marginal product of health ($c > 0$). But also:

(b) if $\beta > 0$ and $\epsilon > 1$

In this case two inconsistencies of the model could result in the same sign prediction as that of the model. Similarly, if the estimated reduced form parameter were found to be positive, it is impossible to determine which of the two hypotheses is violated. Wagstaff (1986) shows that the structural parameter estimates provide a stronger test of the Grossman model predictions because the coefficient of education in the structural demand for medical care equation equals β . Its sign and magnitude can therefore be tested directly.

A second interesting result on education derived by Wagstaff is that exactly the same problems occur when Cropper's (1981) and Muirinen's (1982a) assumption is made that education affects the rate of depreciation δ rather than the efficiency of non-market production. If, for example, the depreciation rate function $\delta(t)$ is assumed to be of the loglinear form suggested by Cropper (1981):

$$\ln \delta(t) = \gamma_0 + \alpha_0 t + \beta_0 E \quad (2.18)$$

then β_0 should be negative if it decreases the use intensity of health by lowering the decay rate. However, the reduced form coefficient of education in the demand for medical care equation then becomes $-\beta_0(\epsilon - 1)$ and its estimation gives rise to identical interpretation problems as described earlier. Wagstaff (1986) therefore concludes that, although conceptually distinct, both approaches to the impact of education cannot be distinguished at the operational level.

The third important point about the effect of education raised by Wagstaff (1986) is related to the pure consumption submodel of the demand for health. Recall from (2.15) that when the production

benefits of health (term (b)) are neglected, the remaining expression for the equilibrium condition contains the Lagrange multiplier λ which is called the marginal utility (or shadow price) of wealth. In the continuous time model formulated by Wagstaff (1986), the analog is $\lambda(0) e^{(\pi-r)t}$ where $\lambda(0)$ is the shadow price of initial assets⁵, r is the interest rate and ρ is the rate of time preference ($r = \rho$ in Grossman (1972a) and Muurinen (1982)). The unobservable $\lambda(0)$ then becomes an important variable in the demand for health function. Muurinen (1982b) and Grossman (1972a) approximated $1/\lambda(0)$ by income, but Wagstaff (1986), using results from lifecycle models of labour supply and consumption (e.g. Heckman, 1976; MaCurdy, 1981), assumes $\lambda(0)$ to be a function of education besides other variables such as life-time wage income, and initial assets. When predicted, respectively postdicted, values of the latter two variables are added to the structural demand function, the effect of education changes from significantly positive to significantly negative! This is not so surprising since education is one of the variables used to estimate the values of the other two variables, and therefore the degree of multicollinearity is raised considerably, but the finding casts doubt on the earlier prediction-confirming positive estimates of the education coefficient. Wagstaff (1986) concludes: 'In view of the fact that 'years of education' tends to be highly correlated with both initial assets and lifetime wages, there can be little doubt that in many of the studies in this field education is acting as a proxy for the real life-time wealth variable $\lambda(0)$ ' (p. 223). As a response to the fact that in the pure consumption version of the model both education and a number of other variables change sign in both the demand for health and some demand for medical care equations, other methods are suggested to overcome the unobservability of $\lambda(0)$ which avoid the multicollinearity problems (Wagstaff, 1986, p. 229). Alternatively, one could question the theoretical model from which the equations have been derived, especially because other crucial model ingredients like the derived demand hypothesis and the use-related depreciation hypothesis are also rejected by the econometric tests.

This last digression serves to illustrate the problems which are encountered when one tries to establish the effects of education in a lifecycle model of the demand for health and medical care. Wagstaff's (1986) rigorous testing of Grossman's (1972a) basic model certainly does not reinforce the confidence in the model as a description of decision-making about medical care use. Perhaps the view of medical care as an investment in health is not very adequate for the bulk of curative medical care consumed. This refers to an early criticism by Usher (1975): 'Most medical expenditure is not a profitable investment unless I happen to be ill' (p. 215). Indeed, the marginal health product of most curative care is zero or negative unless people are sick or disabled. Moreover, even if the Grossman model is more designated to describe preventive health behaviour, it may well be the case that medical inputs represent only a minor share in effective preventive activities compared to general lifestyle. How then could

the decision to demand (curative) medical care be modelled more appropriately? A suggestion was made by Usher (1975): 'An alternative way of modelling the decision to purchase medical care is to suppose that health is a disposition not to be sick too often, that falling ill is a random process dependent on the state of one's health, and that medical expenditure can be either a cost of being sick today or an investment in health tomorrow.' (p. 215). This idea will be one of the subjects of attention in the next section where uncertainty about future health is introduced in the demand models.

Another major concern with the deterministic models described so far is expressed by Evans (1981) as follows: 'A more serious weakness, however, is the common but aggressively counterfactual assumption that consumers possess full information about health production functions' (p. 354). Indeed, the second type of uncertainty, referred to by Arrow (1963) as a crucial characteristic of the demand side in the health care market and leading to the issue of the patient-physician relationship as being one of agency, is completely ignored. In the following section, we will examine the consequences of relaxing the assumptions of complete certainty and full information in a static framework.

2.5. Stochastic models of the demand for health

In order to introduce uncertainty into behavioral models, usually some parameter is assumed to be a random variable with a known probability distribution. Consequently, in the consumer's objective function utility is replaced by expected utility. In this section we will distinguish between two types of uncertainty models, following Arrow's (1963) distinction: uncertainty with respect to health status and uncertainty with respect to the effectiveness of medical care. Both types of uncertainty lead to different derived demands: uncertain health leads to a derived demand for health insurance, and uncertainty as to the quality of the product of medical care leads to a derived demand for information.

2.5.1. Uncertainty with respect to health

Grossman (1972a) made some suggestion as to how uncertainty could be introduced into his model. It could be postulated that the consumer faces a probability distribution of depreciation rates. Using the state-preference approach (see Ehrlich and Becker, 1972) to choice under uncertainty, some results can be reached, none of which, however, tend to alter the major conclusions of his deterministic framework. Generally, consumers might have an incentive to hold 'excess stocks' of health in relatively desirable states in order to avoid 'losses' in more unfavourable 'states of the world'. No formal

derivation is given in Grossman (1972a).

An uncertainty model in which the random losses approach is elaborated in a static framework in order to derive some predictions regarding the demand for health insurance is given by Phelps (1973, 1976). Health H enters the utility function together with a composite good c and is produced by medical care (m) and own time in fixed production coefficients. The stock of health is subject to random losses \tilde{l} having a known distribution $f(\tilde{l})$. The health equation then becomes

$$H = H_0 - \tilde{l} + g(m) \quad (2.19)$$

where H_0 is the endowed level of health, \tilde{l} the random amount of losses (illnesses) and $g(m)$ the health product of medical care. An optimal coinsurance rate and maximum payment amount are derived from maximizing expected utility over all possible states of illness and subject to a budget constraint. In later work (e.g. Keeler, Newhouse and Phelps, 1977), this basic model is extended to a multi-period model in order to examine the effects of deductibles and coinsurance rates on the demand for medical care services.

In all the above mentioned applications of the random losses approach, the uncertainty of health is exogenous, i.e. the probabilities of falling ill are not influenced by the individual's decisions about medical care use. As was mentioned before, this does not seem to be a realistic assumption in the case of preventive activities (such as the use of preventive medicine) because one would hope that these have a negative impact on the probability of future illness occurrence. In order to allow for this endogeneity of the probability distribution of ill health, Phelps (1978) has modified his original model so as to make this distribution a function of preventive inputs. As before, expected utility is defined as the sum of all utility states multiplied by the probability of being in the corresponding state of health. For simplicity, only two states are being considered: the sick state (with probability π) and the healthy state (probability $1-\pi$):

$$E(U) = \pi (c_1, H_1) + (1-\pi) (c_2, H_2) \quad (2.20)$$

where c_1, H_1 and c_2, H_2 refer to the consumption of other goods and health in the healthy and the sick state respectively. The consumers health level in period t is now determined by

$$H_t = H_0 - \tilde{l} + g(P, m) \quad (2.21)$$

with H_0 endowed health, \tilde{l} a stochastic variable which takes the value zero with probability $1-\pi$ and the value l with probability π . Health production g is a function of both the use of preventive medical care (P) and acute medical care (m). However, preventive care is also assumed to have a negative effect on both the illness probability ($\pi_p < 0$), and on length of illness period L ($L_p < 0$), whereas acute medical care m only offsets total sick time: $L_m < 0$. Sick time L

enters the model through the 'full income' budget constraint which also incorporates the time and money prices of preventive and acute care and other goods. Phelps (1978) formulated first-order optimality conditions for this problem but did not explicitly derive comparative static predictions for the demand for preventive and acute medical care. Rather, he assessed the value of preventive medical care, such as direct utility gains, work-loss-time and acute medical care expenditure avoided with the view to draw some implications for private and public health insurance. It is shown, for instance, that under certain conditions and because prevention and acute care are substitutes, extending medical care insurance could reduce the demand for preventive care (see Phelps, 1978, p. 190). He does not discuss the introduction of schooling effects into this framework.

The latter modification is made by Kemna (1985) in the stochastic version of his model of the demand for health and occupational hazards. He makes the probability of falling ill not only a function of the consumption of preventive medical care P , but also of the endowed health level H_0 , of the job hazards z and of the schooling level S :

$$\pi = \pi(H_0, P, z, S) \quad (2.22)$$

with $\pi_z > 0$; π_{H_0} , π_S , $\pi_P < 0$; and $\pi_{ij} = 0$ except $\pi_{PS} \leq 0$.

This last second-order cross derivative is non-positive to reflect Grossman's education-efficiency hypothesis for preventive care: schooling is assumed to reinforce the negative marginal impact of preventive care on the illness probability. The marginal product of curative medical care, however, is assumed to be zero in the healthy state. The distinction between curative and preventive medical care introduces an additional decision variable in the optimization problem. It is demonstrated that maximization of expected utility by a risk averse individual leads to similar comparative static predictions as the deterministic model regarding the effect of education on the demand for curative medical care. Recall from section 2.3 that this effect was ambiguous. The marginal impact of schooling on the demand for preventive medical care can be shown to be positive under certain conditions, which is not surprising given the assumptions. Neither Phelps (1978) nor Kemna (1985), have estimated demand functions for preventive medical care.

Sindelar (1982) uses a very similar two-period model to explain sex differences in the use of preventive medical care. It predicts that individuals with a higher probability of death (or health loss) from causes of death which cannot be deterred by the prior use of medical care will, *ceteris paribus*, use less of that care. The higher risk of violent death for males is estimated to account for about 30% of the observed difference in average expenditures on doctor's care. Coyte (1983) adopts essentially the same assumptions to define the difference between preventative and remedial services in modelling consumer behaviour on the demand side of a market equilibrium

analysis.

Finally, endogenous uncertainty in health has been incorporated in a lifecycle model of health investment by Cropper (1977). She also uses the distinction between the sick and healthy state but in a more dynamic context than the models discussed previously. The health stock is assumed to decay at a constant rate δ over time but depreciation can be prevented by investment in preventive medical care M_t ; i.e.

$$H_t = M_t - \delta H_t.$$

When the health stock H_t falls below a certain critical sickness level H^* no utility is derived from consumption of other goods. This critical level of exposure to illness H^* has a known random probability distribution $f(H)$. The probability of being healthy then becomes

$$F(H) = \int_0^{H_t} f(\bar{H}) d\bar{H} \quad (2.23)$$

As a result, expected utility at time t can be written as

$$EU_t = F(H) U(Y - pM_t) \quad (2.24)$$

where $c_t = Y - pM_t$ is the consumption of other goods when total income Y minus medical expenditure (pM) is spent on c . Maximizing discounted lifetime expected utility, given an initial stock of health and a fixed time of death, an optimal path of investment and health capital is derived. The most interesting prediction from this version of Cropper's (1977) model⁶ is that investment in health is likely to decrease steadily with age after early adulthood. This striking contrast with Grossman's (1972a) result of increasing investment in health over time is, of course, a consequence of the preventive nature of medical care expenditures in this model which yield returns over a shorter remaining period of life at higher age.

Although this model is illustrative of the way in which intertemporal decision-making and the randomness of illness can be, and perhaps should be, combined in a stochastic lifecycle model, it is not immediately useful for our purposes because the role of education or knowledge is not explicitly taken into account. Furthermore, the lack of empirical testing is indicative of the fact that increased mathematical complexity at the theoretical level also increases the difficulties involved in deriving testable predictions. For the time being, it therefore probably remains an elegant conceptual framework which may help to shape thinking about health behaviour as investment.

2.5.2. Uncertainty with respect to the effectiveness of medical care

All models we have discussed so far implicitly assume that consumers of medical services know exactly what medical care can or cannot

contribute to their health. In other words: they possess full information on the nature of their health production functions. Among others, in particular Pauly (1980) has drawn the attention to the inappropriateness of this assumption given the widespread recognition of consumer ignorance in medical matters. He has also suggested how this type of uncertainty could be introduced in the demand for health models. Starting from the simple single period model with utility determined by health H and other goods c , i.e. $U = U(H, c)$, and with the health production

$$H = H_0 + g \cdot M, \quad (2.25)$$

and the budget constraint given by

$$Y = c + pM, \quad (2.26)$$

he assumes the marginal health product of medical care g to be uncertain. So, in this model not H but rather \tilde{g} becomes a random variable with a known (subjective) probability distribution $f(\tilde{g})$. Substituting all constraints into the objective function results in the following expected utility maximization problem:

$$\text{Max } E(U) = \int_G U(Y - pM, H_0 + \tilde{g} \cdot M) f(\tilde{g}) d\tilde{g} \quad (2.27)$$

where G denotes the integration interval for g .

Optimality conditions for this problem require medical care to be consumed so that expected marginal utility is equalized for all possible values of g :

$$\forall_{i,j} : \pi_i U'(g^i) = \pi_j U'(g^j) \quad (2.28)$$

In a second stage, Pauly (1980) adds the possibility for the consumer to obtain additional information on \tilde{g} from a physician. In order to be able to combine the physician's advice of the effectiveness of medical care M in the particular health state H_0 with the individual's own prior belief, it is necessary to assume that patients behave as Bayesian decision-makers. If it is assumed that diagnostic and therapeutic information supplied by the physician also comes in the form of a probability distribution, Bayes' rule provides a mechanism to transform prior knowledge and external advice into a posterior distribution. If $f(\tilde{g})$ is the prior (subjective) probability distribution, $f(\tilde{g}^a)$ is the (unconditional) distribution of the physician's advice, $f(\tilde{g}^a | \tilde{g})$ indicates the conditional distribution (or the likelihood function) of the advice and $f(\tilde{g} | \tilde{g}^a)$ is the posterior distribution, then Bayes' rule can be written as

$$f(\tilde{g} | \tilde{g}^a) = \frac{f(\tilde{g}^a | \tilde{g})}{f(\tilde{g}^a)} f(\tilde{g}) \quad (2.29)$$

with $f(\tilde{g}^a) = \int f(\tilde{g}^a | \tilde{g}) f(\tilde{g}) d\tilde{g}$

Since the ex post (i.e. after receiving advice) decision of the consumer about treatment will depend on $f(\tilde{g} | \tilde{g}^a)$, and not on $f(\tilde{g})$, it is important to know which characteristics of the prior distribution

and of the adjustment factor (the ratio of the conditional and unconditional distribution of the advice \tilde{g}^a) determine how much the posterior deviates from the prior. Pauly (1980) shows what happens in two extreme degenerate cases (complete certainty and complete ignorance of the patient) but does not derive any general comparative static results: "One can say that the response to new information of the individual's use of care is likely to be different for persons with different prior beliefs or prior stocks of information. But it does not appear that one can make any general a priori conjectures about the direction of this relationship" (p. 47). However, he does show that in the bivariate case, when the distribution of \tilde{g} is characterized by only one (prior) probability π , the effect of the physician's advice \tilde{g}^a on the change in this π is smaller the larger is $|\pi - \frac{1}{2}|$. The proposition that the more informed are less responsive to information than the ignorant (with π close to $\frac{1}{2}$), becomes the basis of his empirical work. We will defer the discussion of this empirical analysis to Chapter 4.

The crucial thing to note from Pauly's (1980) suggestion is that the Bayesian approach provides the missing link between on the one hand the demand models based on sovereign, fully informed, rational consumers discussed in previous sections, and, on the other hand, the so-called 'need' models discussed in various health economics textbooks (e.g. Cullis and West (1979), Newhouse (1978)) which imply that the decision about how much medical care to use is entirely determined by the medical expert. These two opposed viewpoints become in the Bayesian context just two extreme cases of the general principle that decisions are based on a mixture of own and expert information. As a consequence, patients do not only demand treatment but also diagnostic and therapeutic information which is often obtained from the same provider. The provision of information then becomes the vehicle through which suppliers can manipulate demand. The accuracy of information also becomes a new decision variable in the models of physician behaviour which are discussed by Pauly (1980) at a rather global conceptual level. In sum, the main message of Pauly's (1980) reasoning is that the demand for and supply of information should form an integral part of any attempt to model medical care transactions. Although his analysis is primarily conceptual, it clearly indicates that the Bayesian approach has a lot of potential to offer in deriving the demand for information and explaining its effect on the demand for medical care.

The Bayesian approach to modelling information asymmetry is also adopted by Coyle (1984). In order to be able to explain the 'stylized fact' of demand creation in health care, he relaxes the full information assumption in his equilibrium analysis of the market for medical services based on hedonic price theory. Consumers are assumed to be uncertain about the effectiveness of medical care in reducing the length of an illness episode. Physicians can raise demand for their services by overstating the value of medical care but this ability is

constrained by the patient's possibilities for monitoring physicians. They only engage in monitoring whenever the marginal benefit from monitoring exceeds its marginal cost. The predictions derived from this model are basically the same as in Pauly (1980) but no empirical testing is pursued.

2.6. Conclusions and implications for research strategy

The survey of literature presented in this chapter was set up in such a way as to indicate the development in the theory of the demand for medical care. Basically, four major characteristics could be identified: (1) the derivation of the demand for medical care from the more fundamental demand for health, (2) the investment character of medical care use when health is regarded as a durable stock of human capital, (3) the random nature of health losses which causes a derived demand for insurance and allows for a distinction between preventive and curative care, and (4) the imperfect information on the effectiveness of medical care which induces a derived demand for diagnostic and therapeutic information. In table 2.1 most of the models discussed in this chapter are ordered according to three dividing lines. This survey is not exhaustive but it does present a useful description of what modifications are necessary in order to extend the basic textbook model of consumer demand.

Table 2.1. Survey of formal models of the demand for health and medical care

DETERMINISTIC		STOCHASTIC	
		Health losses	Effect of medical care
STATIC	Phelps & Newhouse (1974) Acton (1975) Kemna (1985)	Phelps (1976, 1978) Sindelar (1982) Kemna (1985) Coyte (1983) Keeler et al. (1977)	Pauly (1980) Coyte (1984)
DYNAMIC	Grossman (1972a, b) Muurinen (1982a) Cropper (1981) Wagstaff (1986a)	Cropper (1977) Hey & Patel (1983) Ippolito (1981)	

What have we learned from this overview of demand models that is useful for the objectives of this study? First, the relationship between the demand for medical care and the consumer's stock of

knowledge cannot be investigated in a meaningful way without also incorporating the more fundamental demand for health. The main reason for this is that in most models the influence of education (as a proxy for knowledge and skill) is channelled through the individual's production function of health. It is then assumed to operate at the purely technical level of health production as an efficiency-improving factor. This means that, in circumstances when medical care is potentially beneficial to one's health, the better schooled get more health out of the same use or, alternatively, need less care to achieve the same health improvement. This assumption is standard practice in household production theory and yields - under certain conditions - testable predictions regarding the effect of education on the demand for health and medical care in the structural form equations of the model (cfr. Wagstaff, 1986a). However, it overlooks the basic criticism that, even for very informed consumers, the nature of the health production function remains highly uncertain. If, as argued by Pauly (1978, 1980), the demand for medical care is largely a demand for information in order to reduce this uncertainty, the predictions regarding the effect of education may well differ from Grossman's (1972a). Especially because also empirical tests of the education-efficiency hypothesis were not convincing, it is worth pursuing a different approach. This will be done in Chapter 3 and 4 where it will be assumed that education induces a reduction in the uncertainty around - rather than an increase in the magnitude of - the marginal effectiveness of medical care. In addition, by distinguishing between general knowledge (e.g. education level) and specific knowledge (e.g. medical), it may be possible to discriminate both approaches empirically.

An alternative route to introducing education into the demand for health models was advocated by Muurinen (1982a). She suggested that the direct, and permanent, effect of education as an environmental variable on the use-intensity of the health stock might be much more important. In her model, education is not seen as an intervening variable in household production processes but rather as having a beneficial influence on decisions about the allocation of time and resources to production and consumption. Although the distinction between both approaches is basically a matter of convention because they lead to similar predictions about the effect of education on the demand for health and medical care, the 'allocative' health benefits of education indeed intuitively seem to be more important than the purely technical benefits. In the second part of this thesis, we will attempt to test for these allocative benefits directly by analyzing health changes over time using panel data. So, rather than estimating structural or reduced form demand equations from cross-section data, we will concentrate on the basic identity which describes the net investment in health (equation 2.8 in Section 2.4).

From the above, it has become clear that we have split the remainder of this thesis into two parts. The first part explores the relation-

ship between knowledge and the demand for medical care using a static but stochastic theoretical model. In the second part, the more direct linkage between education and health is examined in an explicitly dynamic but deterministic context. Ideally, of course, one would prefer to tackle all issues simultaneously by means of a comprehensive theoretical model of lifecycle behaviour incorporating all sources of uncertainty. However, the literature review has shown that the most complex models were not necessarily also the most fruitful in producing testable predictions. Stochastic lifecycle models found in the literature on health behaviour (e.g. Cropper, 1977; Ippolito, 1981; Hey and Patel, 1983) were not accompanied by any empirical analysis. Some support for the approach adopted in this thesis can be found in the following quotation from Muurinen (1982) in the appraisal of her own economic model of health behaviour: "The generality of the model and the extent of problems it can analyse are its main advantages. However, in order to analyse specific problems in greater detail, it is necessary to concentrate on particular aspects of the model and to develop these submodels further, possibly at the expense of major simplifications in the rest of the framework." (p. 22).

Notes

1. Real full wealth w^* is defined as full income y plus healthy time valued at its shadow price and divided by an average weighted price level (see Kemna, 1985, p. 39).
2. The ambiguity is caused by the fact that schooling raises both the income and the healthy time component of full wealth and either increases or decreases the shadow price of health (see Kemna (1985), Appendices B and C).
3. A formulation of essentially Grossman's model in a single period static framework can be found in Wagstaff (1986b).
4. A similar procedure was also adopted by Cropper (1981) in a pure investment model of the demand for preventive health care without explicit reference to the concept of use intensity.
5. In terms of optimal control theory, $\lambda(0)$ is the value at $t = 0$ of the co-state variable of the dynamic asset accumulation constraint (see e.g. Dorfman (1969) or Kamien and Schwartz (1981) for this terminology).
6. In the same article, she also analyzes the dynamics of two other variants of this model, one with time of death endogenous and one in which the trade-off between higher wages and occupational hazards is examined.

**Part I. CONSUMER INFORMATION AND THE DEMAND
FOR MEDICAL CARE**

Chapter 3. The demand for information on the quality of medical care

"Frequently, the only way a person can know whether he needs to see a physician, is to see a physician" (Fuchs, 1974, p. 147)

3.1. Introduction

The literature survey of Chapter 2 has brought out that one potentially interesting approach to the relationship between a consumer's knowledge or information and his use of medical care has been left somewhat underdeveloped. The idea to view patients as Bayesian decision makers has been suggested by Pauly (1980) but he did not explore its consequences in great detail. In this chapter, we attempt to explore the issue further. First, we will briefly point out how the asymmetry of information between demanders and suppliers has generated some special features of the market for medical care. A distinction will be made between intrinsic special characteristics of health care and characteristics which are largely responses (e.g. government intervention) to these peculiarities. Secondly, we present a formal model of consumer behaviour which is essentially an application of Kihlstrom's (1974) model of the demand for information about uncertain product quality to the demand for medical care. Some of its predictions will be tested in Chapter 4. Finally, before we draw some conclusions, section 3.4 illustrates Pauly's (1980) information-manipulation hypothesis in the Bayesian context.

3.2. Asymmetry of information in medical care

Several authors have reviewed the peculiarities in the health care sector in order to derive welfare economic implications for the functioning of the market mechanism. Arrow (1963) drew the attention to the pervasiveness of uncertainty: "I will hold that virtually all

the special features of this industry, in fact, stem from the prevalence of uncertainty" (p 19). He distinguished between two types of uncertainty (cfr. section 2.5). Type one is the uncertainty with respect to the occurrence of illness. Insurance markets have developed to cover the financial losses associated with health risks. This did not occur for the (type 2) uncertainty about the effects of medical treatment. Ideally, such an insurance against a failure to benefit from medical care would involve the transfer of the risk from the patient to the physician. We did not see the emergence of this sort of insurance contracts, according to Culyer (1971), "because of the enormous costs of discovering whether treatment had been successful" (p. 58). The informational imbalance between patient and physician makes enforcement of standards of care very difficult. Therefore, other mechanisms have been used instead to reduce the uncertainty to the patient. In order to protect ignorant consumers, governments have created barriers to entry in the medical profession. Exclusive licensure of medical doctors is assumed to ban low quality providers from the market. The special trust relationship between patients and their physicians, guided by an ethical code which also prohibits commercial advertising, and the requirement to obtain a prescription for ethical drugs, are other examples of institutional arrangements of this industry which are in fact a consequence of the basic uncertainty about product quality.

But patients can also engage themselves actively in informational activities: "When there is uncertainty, information or knowledge becomes a commodity. Information, in the form of skilled care, is precisely what is being bought from most physicians, and indeed from most professionals" (Arrow, 1963, p. 18). Instead of considering medical care consumption merely as a delegation of decision-making to an expert-agent, it can also be viewed as a transfer of information. Pauly (1978) argues that most medical care contacts indeed consist for a large part of provision of diagnostic and therapeutic information. Especially primary care providers spend a great deal of their time informing patients about their health status and about the health outcomes to be expected from different therapies. This advice has to be distinguished from the role played by advertising in other markets because the provision of information and services is integrated in one joint product.

The natural question which arises then is whether neo-classical demand theory can handle this problem. Arrow (1963) is sceptical: "the demand for information is difficult to discuss in the rational terms usually employed.... The elusive character of information as a commodity suggests that it departs considerably from the usual marketability assumptions about commodities (p. 18)". Other authors do not see any obstacles at all; compare Telser (1981): "the demand for drugs is related to the demand for the more general good health in the same way that the demand for automobile parts is related to the demand for automobiles". In fact, the solution to the problem may be somewhere in

between these two positions: the neo-classical paradigm can be extended to allow for uncertainty about product quality.

Kihlstrom (1974a, b) has adopted a Bayesian approach to consumer decision-making to develop a general theory of the demand for information. He showed that, as soon as product quality is uncertain, information or knowledge becomes a commodity which is demanded by consumers, not because of any desirable attributes it possesses in itself, but because it helps them to make 'better' purchases of the uncertain product. This may give rise to 'unusual' demand behaviour in the neo-classical sense. Ignoring this in models of the demand for medical care could mean a serious misspecification because: "very few industries could be named where the consumer is so dependent upon the producer for information concerning the quality of the product" (Fuchs, 1966, p. 137). In the next section, we will present Kihlstrom's (1974a) model and investigate its applicability to the demand for medical care.

3.3. A Bayesian approach to the derived demand for information

3.3.1. The Kihlstrom model of Bayesian consumer behaviour

In this section, we describe a static model of consumer behaviour under uncertainty which resembles the one discussed in Pauly (1980). It draws heavily on Kihlstrom (1974a) who discussed the demand for information about the quality (i.e. the amount of carbohydrates provided) of cereal. Consumers are assumed to derive utility from two commodities: health H and some other composite good c . The latter can be bought in the market at a constant price p_c , but health is not for sale. Instead, health improvements can be obtained from using medical care m bought at the unit price p_m . The health production function has the following simple form:

$$H = H_0 + b m \quad (3.1)$$

where m = amount of medical care utilization

H_0 = endowed health level at the beginning of the period

and b = marginal health product of medical care

Note that in this framework, in contrast to the dynamic health models described in Chapter 2, H_0 is predetermined. The consumer may lack both perfect diagnostic information (what is H_0 ?) and therapeutic information (what is b ?). We will assume that he knows what his initial health level H_0 is, but that he is uncertain about the marginal effect of medical care, i.e. b becomes a random variable \tilde{b} .¹ Of course, the effect of medical treatment b does have some relationship with the initial health condition H_0 . It can alternatively be written as follows:

$$b = \frac{H - H_0}{m} = \frac{\Delta H}{m} \quad (3.2)$$

This expression indicates clearly that b measures the marginal health increment per unit of medical care consumed. It was pointed out in the previous chapter that the marginal product of preventive and curative medical care can be distinguished if health is assumed to be subject to random losses. Preventive care reduces the probability of a loss (illness episode) but curative care is only effective in reducing sick time if a loss has occurred. Strictly speaking, this implies that there is no (or at least very little) uncertainty about the effectiveness of curative medical treatment in the healthy state because it is perceived to be zero. Without this uncertainty, there can be no demand for information in the healthy state. This means that for curative care - which represents the major part of health care delivery - the demand for information can only arise in periods with some (perceived) health losses.

Although the consumer does not know the value of b with certainty, he does feel able to attach probabilities to the various possible states of the world. This prior belief is summarized by a normal probability distribution with mean b_0 and variance $\frac{1}{\phi}$.²

$$\text{So let } \tilde{b} \sim N(b_0, \frac{1}{\phi}) = f(\tilde{b} | b_0, \frac{1}{\phi}) \quad (3.3)$$

In the consumer's subjective beliefs, the parameter b_0 is his best guess of the marginal effectiveness of medical care for his health status. The parameter ϕ expresses the confidence in his estimate. When he is unfamiliar with medical care, ϕ will be very small and thus the variance of \tilde{b} around b_0 very large. Without any extra information available to him, the consumer maximizes expected utility subject to his budget constraint, i.e.:

$$\max_{m, z} E(U) = \int_{-\infty}^{+\infty} U(z, H_0 + b m) f(b | b_0, \frac{1}{\phi}) db \quad (3.4)$$

$$\text{s.t. } y = p_z z + p_m m$$

where: y = income

p_m = unit price of medical care m

p_z = unit price of composite good z

This maximization problem leads to a unique solution (z^*, m^*) if U is strictly concave in z and H . The presence of uncertainty, implied by $\frac{1}{\phi} > 0$, makes additional information on b valuable to the consumer.

This information can be acquired from several sources; e.g.:

1. personal observations on b by own experience,
2. observations taken from the experiences of others, or
3. purchased expert medical advice.

Regardless of the information source, the acquisition is usually costly, either in terms of time or money. These costs have to be weighted against the benefits of making more informed and therefore

'better' purchases. In this way the demand for information about medical care is derived from the uncertainty of its effect on the consumer's health. Bayesian decision theory provides a method to incorporate this (sample) information into the decision-making process. In Kihlstrom (1974a), the consumer can obtain sample information about the quality of cereal from a research laboratory. It is assumed that this information comes only in the form of a "summary" given by the likelihood function $f(x | b, \frac{1}{\theta})$. This is the conditional density of a random variable x_θ given b where

$$\tilde{x}_\theta = \frac{1}{\theta} \sum_{i=1}^{\theta} \tilde{y}_i \quad \text{with } \tilde{x}_\theta \sim N(b, \frac{1}{\theta}) \quad (3.5)$$

\tilde{x}_θ is in fact the sample mean of θ independent observations $\tilde{y}_1, \dots, \tilde{y}_\theta$. Each observation y_i is itself a random variable, normally distributed with mean b and variance 1. In that case, the sample mean is said to be a sufficient statistic, which means that observing \tilde{x}_θ is equivalent to observing $\tilde{y}_1, \dots, \tilde{y}_\theta$ (see e.g. Degroot, 1970, p. 155). By using sample means of normally distributed observations as the units in which to acquire information, θ provides a measure of the amount of information units provided by observing \tilde{x}_θ .³ A higher θ implies more information. If θ is large, the consumer can be fairly confident that observing x is much like observing b .

It is further hypothesized that the consumer can choose the amount θ of information to be obtained at a cost $q\theta$ where q is the price of one observation \tilde{y}_1 . In the present context and at this stage of the discussion, it is useful to think of \tilde{y}_1 as one physician advice. This means that in order to acquire more information, more opinions have to be obtained.

In deciding how much information to obtain (which sample to observe) the consumer will have to weight the costs of purchase against its value in terms of utility gains. How much utility can be gained from better decisions as a consequence of better information? It is useful to consider first the extreme but simpler case in which perfect information can be obtained. Recall from (3.4) that $E(U(z^*, m^*))$ is the utility maximum for the individual when he only has his subjective prior distribution of b to rely on. Now suppose that he is able to observe what the value of \tilde{b} will be. His optimal decision is then to choose m and c , given $\tilde{b} = b$, so that $U(z, m)$ is maximized, which gives him $U(c(b), m(b))$. This could be called the 'ex post utility', i.e. after receiving perfect information. To arrive at the expected utility (ex ante) with perfect information, this procedure has to be repeated for each possible b . In fact, the ex post utility of observing $\tilde{b} = b$ has to be weighted by its probabilities of occurring given by the a priori density function $f(b | b_0, \frac{1}{\phi})$. Expected utility of having perfect information therefore equals

$$E[U(z(b), m(b))] = \int_{-\infty}^{+\infty} U(z(b), H_0 + b m(b)) f(b | b_0, \frac{1}{\phi}) db \quad (3.6)$$

It can be shown that the expected value of receiving (perfect) information is non-negative, i.e.⁴

$$E[U(z(b), m(b))] - E(U(z^*, m^*)) \geq 0 \quad (3.7)$$

In other words: acquiring perfect information costlessly cannot make the consumer worse off.

In general, however, he will not be able to obtain perfect information. Instead, imperfect statistical information may be acquired in the form of x , the value taken by \tilde{x}_θ , before a choice is made about how much z and m to purchase. This imperfect information does not consist of an observation of the actual b ex post, but of the sample information provided by \tilde{x}_θ .

Again, in order to derive an expression for the ex ante expected utility under conditions of imperfect information, a two step procedure has to be followed. First, we compute the utility of θ units of information obtained by observing that \tilde{x}_θ equals x . This is called the expected a posteriori utility when x has been observed. We can then no longer use the a priori density of b for the weighting procedure, but have to weight with the revised probabilities of b determined by the posterior density. Since the consumer is a Bayesian decision-maker, and both the prior density of

$$\tilde{b} \ (\tilde{b} \sim N(0, \frac{1}{\phi}))$$

and the conditional density of \tilde{x}_θ given b ($\tilde{x}_\theta \sim N(b, \frac{1}{\theta})$) are normally distributed, it can be shown (cfr. e.g. Degroot, 1970) that the posterior will also have a normal distribution with mean $b_{x, \theta}$ and variance σ_{θ}^2 where

$$b_{x, \theta} = \frac{\theta}{\theta + \phi} x + \frac{\phi}{\theta + \phi} b_0 \quad (3.8)$$

$$\text{and } \sigma_{\theta}^2 = \frac{1}{\theta + \phi} \quad (3.9)$$

The posterior mean of b' is a weighted average of the prior expectation b_0 and the observed sample mean x . The weights represent the balance of the confidence in the prior belief (ϕ) and the 'accuracy' of the sample information (θ). The relative magnitude of θ and ϕ determines whether the posterior will be closer to the prior or to the sample mean.

Note also that the posterior variance is always lower than the prior variance. This implies that adding the sample information always results in an increase of the prior confidence. Now the expected posterior utility when x has been observed can be written as

$$EU_{x, \theta}(z, m) = \int_{-\infty}^{+\infty} U(z, H_0 + b m) f(b | b_{x, \theta}, \sigma_{\theta}^2) db \quad (3.10)$$

If information can be bought at a constant unit price q , the consumer's budget is accordingly to be reduced to $y - q\theta$ when θ units of information are acquired. The utility of these θ units of inform-

ation is now to be weighted with the probability of observing x . For this purpose, we must use the **marginal** (or unconditional) **density of x** . This can be derived by integrating the product of the conditional density of \tilde{x}_θ given b (i.e. $f(x|b, \frac{1}{\theta})$), and the marginal density of b (i.e. $f(b|b_0, \frac{1}{\phi})$) over the entire interval $(-\infty, +\infty)$ of b . Thus, the marginal distribution of x_θ , with mean b_0 and variance $\frac{1}{\theta} + \frac{1}{\phi}$, becomes

$$f(x|b_0, \frac{1}{\theta} + \frac{1}{\phi}) = \int_{-\infty}^{+\infty} f(x|b, \frac{1}{\theta}) f(b|b_0, \frac{1}{\phi}) db \quad (3.11)$$

The unconditional probabilities of x represented by this distribution are the relevant weights to be used in computing the ex ante expected utility of θ units of information (that is, before we know which x is going to be observed) by means of the following expression:

$$H(\theta) = \int_{-\infty}^{+\infty} EU_{x, \theta}(z(x, \theta), m(x, \theta)) f(x|b_0, \frac{1}{\theta} + \frac{1}{\phi}) dx \quad (3.12)$$

The resulting consumer's demand for information is the value of θ which maximizes $H(\theta)$ subject to the budget constraint $y = p_z z + p_m m + q\theta$. It is clear from this expression that the maximum value of $H(\theta)$ will be a function of the 'standard' determinants like relative prices p_z and p_m and income y , but also of the parameters b_0 and θ which characterize the consumer's a priori expectations. It is through these parameters that consumer knowledge about medical matters will influence the demand for diagnostic and therapeutic information.

3.3.2. Comparative static predictions

The basic model described in the previous section has been modified and extended by Kihlstrom and others in various ways. First of all, there is a more general version, outlined in Kihlstrom (1974b), which extends the analysis beyond the two-goods case and the assumption of normal densities. Some general theorems can be derived which are analogous to the neo-classical theory of consumer demand - stated in terms of the Slutsky matrix of substitution terms. The analysis does not, however, provide very specific and empirically verifiable predictions about the effects of prior information on demand.

When, in addition to strict concavity, an additional assumption is being made about the functional form of the utility function, stronger results can be obtained. Assuming the following utility function with constant elasticity of substitution one-half

$$U(c, H) = -\frac{\alpha}{c} - \frac{\beta}{H} \quad (3.13)$$

where $\alpha, \beta > 0$, Kihlstrom (1974a) is able to derive a number of interesting comparative static predictions about the demand for information.⁵ We restrict ourselves to a discussion of the most important results here and refer to Kihlstrom (1974a) for the mathematical derivations. The propositions are:

$$1. \frac{\partial \hat{\theta}}{\partial y} > 0 \text{ and } \frac{\partial \hat{\theta}}{\partial q} < 0$$

When information demand is positive ($\hat{\theta} > 0$), it increases with income and falls when the price of information rises.

$$2. \frac{\partial \hat{\theta}}{\partial \phi} < 0$$

The amount of information demanded falls with increasing confidence in the a priori expectations (or familiarity with the product).

3. Information demand is highest for goods for which the expected quality level is neither very high nor very low. This is essentially because with b_0 very large or near zero, the consumption decisions for z and m are relatively independent of the observation of x .

4. No demand for information will arise when:

- a. the uncertain product is either very expensive or very cheap relative to other products
- b. the consumer has very skewed preferences in favour of one commodity
- c. the uncertain product is of either very high or very low quality.

Proposition 1 states that information is in neo-classical terms non-Giffen and non-inferior and proposition 2 is a logical consequence of the way in which uncertainty is defined. Though less straightforward than the previous ones, proposition 3 is also consistent with intuitive reasoning that additional quality information is most valuable when expectations are not extremely optimistic or pessimistic but rather intermediate. A more surprising result is that, not only for very cheap but also for very expensive uncertain products, no quality information will be demanded (unless this means that very expensive products will never be bought either). It seems less relevant for health care where typically very expensive treatment will almost always be covered. It does question the rationale of the legal requirement to buy information (in the form of a prescription) for a number of relatively cheap drugs. 4b is also interesting in the case of medical care. When a serious illness causes preferences to become very skewed in favour of health, again demand for information (relative to demand for treatment) will be low. 4c is simply the inverse of proposition 3. The first three propositions will be tested in the empirical analysis presented in Chapter 4.

The basic Kihlstrom model described above has been modified in later work in order to enable further investigation of other aspects of decision making about the acquisition of information. In section 3.3.1 it was assumed that additional information could only be bought at a constant unit price in the market for information. Information was viewed as the summary of the results of a number of experiments comparable to the sample information used in Bayesian statistical decision theory. Intertemporal effects of information demand were not considered because sequential sampling was explicitly ruled out. In

other words, the consumer could only decide whether or not to buy the information (or message) \tilde{x}_θ as a whole at price q_θ in that period. It was not possible to observe some of the θ observations (e.g. $\theta' < \theta$) and then to decide whether or not to observe the remaining observations ($\theta - \theta'$). However, obviously the information acquired in one period carries over to the next period where the prior information becomes the posterior information of the previous one. This intertemporal extension was modelled explicitly by Grossman, Kihlstrom and Mirman (1977). In their model it is assumed that information is not bought from experts (like medical doctors) but accumulates from one's own previous experience with the product of uncertain quality. Simply buying and consuming medical care also provides observations on its effectiveness which 'update' the prior expectations for future decisions. In order to incorporate such production of information through experimentation in the model of consumer behaviour, the utility function has to be extended to include these 'learning by doing' effects. Bayes' rule again provides the updating mechanism. It is shown that the possibility of learning from experience induces experimental consumption behaviour: consumption is higher than the optimal level without learning opportunities. The cost of the information is the utility foregone in the current period (by overconsumption) and the benefits are the utility gains from 'better' future purchases.

Although we will not be concerned with the experimental aspects of medical care as an 'experience good' (see e.g. Nelson, 1970) in what follows, the results of the intertemporal approach do offer some insight as to where the prior expectations come from. This will prove useful in the next section.

Finally, for completeness one last result concerning the determinants of the demand for product quality information deserves mentioning. Freixas and Kihlstrom (1984) have examined theoretically the effect of attitudes toward risk using a very similar model to the one presented in section 3.3. The remarkable conclusion is reached that quality information demand declines with increasing risk aversion.⁶ This may seem counter-intuitive because one would expect more risk-averse decision makers to have a stronger preference for reductions in uncertainty. But the authors point out that the uncertainty reduction only occurs ex post, i.e. only after the informative message has been received: "Thus, ex ante, the returns to information are uncertain, and more risk averse buyers should be less willing to accept the risks associated with its acquisition" (Freixas and Kihlstrom, 1984, p. 93). Unfortunately, due to data limitations, the prediction on risk aversion could not be tested in our empirical analysis. How we will attempt to test some of the other comparative static results in the context of the demand for (primary) health care is described in the next section.

3.4. From theoretical to empirical model

Before presenting our empirical work, it is worthwhile to discuss some of the problems which arise when one wants to test some of the predictions of the Kihlstrom model. It seems as if the practical difficulties encountered have put off researchers because we could not find any examples of empirical studies testing the theory of the demand for product quality information.

For ease of reference, we have summarized the sign predictions of the model in table 3.1. Some of the theoretical variables included in this table need some further operationalization in the specific context of medical care demand. Without going into details about the data and measurement problems (which are deferred to the next chapter), we will discuss at the conceptual level: (a) the distinction between the demand for the product and for information about its quality; (b) the interpretation of the demand determining factors; and (c) the incorporation of information-manipulation effects. Because the empirical analysis is based on Belgian health care utilization data, the discussion is written with the institutional context of Belgian health care organization in mind.

Table 3.1. Sign predictions of the static Kihlstrom (1974a) model

	Income	Prices	Quality expectations		Preferences	
	y	q	$\frac{p_m}{p_z}$	ϕ	b_0	$\frac{\alpha}{\beta}$
Demand for information $\hat{\theta}$	+	-	nl	-	nl	nl

(nl) indicates a non-linear effect: information demand becomes zero at the extremes and is positive in between

3.4.1. Joint product problems

As we already pointed out in section 3.2, a peculiar characteristic of the health care market is that physicians to a large extent supply both information and medical services. The 'informative-prescriptive' component of the care provided may be difficult to separate from the 'active-therapeutic' component. In Kihlstrom's (1974a) application, the cereal was not purchased from the research laboratory from which the sample information was bought. A parallel can be drawn with the

consumption of drugs. In Belgium and in many other countries, drugs can only be bought from the pharmacist whereas the information or advice about the drugs is mostly obtained from a physician. It seems logical then to use the two most frequent acts of medical consumption, i.e. the general practitioner visits and the use of medicines as empirical counterparts of the demand for information (θ) and the demand for medical care (m).

There are some drawbacks, however, in this approach. First, although drug prescribing and other advice constitutes by far the largest share of activities in Belgian general practice, some of the contacts result in active treatment by the GP. By considering only the GP contacts which occurred at the initiative of the patient it may be expected that at least the initial reason for the contact was to seek advice. Another snag has to do with the institutional arrangements. Most drugs cannot be bought in a pharmacy without a prescription from a medical professional. This legal requirement may artificially raise the demand for information among informed consumers who perhaps do not need the advice. Because the prescription requirement also considerably raises the (time and money) cost of these ethical drugs, it may induce better informed consumers to substitute them to some extent by over-the-counter drugs.

3.4.2. Income, prices, expectations and preferences

Some of the exogenous variables listed in table 3.1 need some further specification with respect to their application to the demand for therapeutic information. The effects of income y and the relative price ratio P_m/P_z have to be interpreted in the context of the Belgian public health insurance system which is described in Chapter 4, section 3.2. Extensive insurance coverage of medical services has caused what Newhouse (1979) calls "the erosion of the medical marketplace". It implies that the traditional market model of arm's length transactions guided by relative price signals does not apply very well to the patient-physician contacts. All patients pay the same fixed deductible amount for a physician visit or for the purchase of a prescription drug: P_m will therefore show no variation in cross section studies and money price elasticities of demand can therefore only be obtained from time series analysis (see e.g. Van Doorslaer (1984) or Carrin and Van Dael (1984)). The only price of medical care which varies between individuals is the price of travel, waiting and treatment time.

This does not apply to q , the price of information in the theoretical model. In the cereal example, the consumer could choose the size of the sample (θ) from which he wanted to observe the sample mean (x) at a cost of $q\theta$. In the case of a patient with a health problem, the decision to be taken is more often about whether or not to seek professional advice, rather than about how much advice to obtain. Now,

if consumers differ in their perception of the accuracy of the advice (which is θ if $1/\theta$ is the variance of the sample mean), then the implicit price of information also varies. If p_v is the (constant) out-of-pocket price of a GP visit and θ_p is the perceived accuracy of the advice, then the implicit price of information $q = p_v/\theta_p$ varies across individuals according to their perceptions of the informative content of a G.P. visit.⁷ Variation in the perceived accuracy of GP's thus generates variation in the perceived implicit price of information. The consumer's estimate of θ_p is derived from own previous experience with his physician's advice and the information acquired from the experience of others. When faced with some health problem and the decision whether or not to seek professional medical advice, a common practice is to discuss this with friends and relatives in order to obtain "information about the accuracy of information". Especially primary care physician's services can be viewed as a "reputation good" (Pauly and Satterthwaite, 1981) for which perceived accuracy of diagnostic and therapeutic information is an important determinant of perceived quality. In section 3.4.3 it is shown that the accuracy perception not only increases the demand for information by lowering its price, but also influences the responsiveness to the advice when it is received. Consequently, perceived accuracy of advice θ_p becomes an explanatory variable in the demand function for information through its impact on q .

The a priori expectations, summarized by the parameters b_0 and ϕ are also not directly observable. The confidence in the a priori estimate of the effectiveness of medical care b will in general increase with knowledge about health and medical care. More knowledgeable consumers are therefore assumed to demand less information, ceteris paribus. The relationship between b_0 and medical knowledge is not so straightforward. Whereas it is plausible to hypothesize that knowledge reduces the variance of the prior distribution it is not clear whether it raises or lowers the average expectations about medical care effectiveness. What can be seen from equation (3.2) in section 3.3.1 is that it will be negatively related to perceived initial health status H_0 . This means that the demand for therapeutic information only becomes positive when b_0 has exceeded some threshold level, or similarly when H_0 goes below some threshold level (cfr. propositions 3 and 4c in section 3.3.2).

Finally, the "relative taste for health" prediction stated in proposition 4b will not be tested in our empirical analysis. The emphasis will be on estimating the effects of the two components of information, ϕ and θ_p . In the next section we will illustrate how they interact in Bayesian decision analysis.

3.4.3. Effects of information-manipulation

Apart from generating predictions on the factors which determine the

demand for information, this behavioural model also yields insight into individuals' responsiveness to such information once they have acquired it. This is not unimportant in health care because of the alleged possibilities for physicians to manipulate the demand for their own services by lowering the accuracy of their advice (Pauly, 1980). According to this information-manipulation theory, the ability to create supplier-induced demand is constrained by two characteristics of the demanders: the confidence in one's own subjective beliefs and the perception of the accuracy of the advice.

Recall that the physician's advice changes the prior distribution summarizing consumer beliefs into a posterior distribution according to Bayes' formula:

$$f(b|b^A) = \frac{f(b^A|b)}{f(b^A)} f(b) \quad (3.13)$$

$$\text{with } f(b^A) = \int_R f(b^A|b) f(b) db$$

where $f(b)$ = subjective prior distribution of b
 $f(b^A|b)$ = likelihood function of the doctor advice b^A
 $f(b^A)$ = unconditional distribution of b^A
 $f(b|b^A)$ = posterior distribution of b
 R = the integration interval of the variable b

Since the ultimate consumer decision about treatment will depend on $f(b|b^A)$, not on $f(b)$, it is important to know which characteristics of the prior distribution and the likelihood function determine how much the posterior deviates from the prior. In general, these deviations will be larger, the more diffuse the prior, and the less diffuse (and therefore more informative) the likelihood function.

To illustrate this general proposition and to enhance intuitive understanding, it is useful to consider some numerical examples of four degenerate situations. Let the patient have a subjective discrete prior distribution function $f(b)$ over 2 possible effects, b_1 and b_2 , of a particular treatment. His beliefs about the accuracy of his doctor's advice on these outcomes are represented by the likelihood function $f(b^A|b)$.

(a) Consumer certainty

This situation is characterized by a variance of the prior distribution equal to zero. One could think of a consumer who has had the same symptoms before and knows which treatment worked and which did not. On the other hand, he perceives the accuracy of his doctor's advice to be less than perfect. For instance, if the actual outcome of the treatment is b_1 , there is only 60% chance that the doctor had predicted it. In 80% of the cases when b_2 was the outcome, it was also predicted to be so by the doctor.

In tabular form, this looks as follows.

prior distribution conditional distribution (likelihood)

	$f(b)$		$f(b^A b)$	b_1	b_2
b_1	1	b_1	.6	.4	
b_2	0	b_2	.2	.8	

In order to obtain the posterior, the unconditional probabilities of b^A are determined by multiplying each conditional probability of b^A with the corresponding prior probability of b . The posterior probabilities are then given by the following Bayes' formula in the discrete case (summing over $i = 1, 2$ instead of integrating over the interval of f).

$$P(b_1|b^A) = \frac{P(b^A|b_1) \cdot P(b_1)}{\sum_{i=1}^n P(b^A|b_i) \cdot P(b_i)} \quad (n = 2) \quad (3.14)$$

unconditional distribution $f(b^A)$ posterior distribution

$f(b)$	$f(b^A b)$	b_1	b_2		$f(b b^A)$	b_1	b_2
b_1		.8	.2		b_1	1	1
b_2		0	0		b_2	0	0
$f(b^A)$.8	.2				

It is clear that in this case the doctor advice has no effect on the decision of the patient.

(b) Consumer ignorance

The other extreme is when the consumer/patient has no idea about the probabilistic outcome of the treatment alternatives. They are then both equally likely to him. After having received the therapeutic advice, these probabilities are revised into a more informative (less uncertain) posterior distribution.

	$f(b)$	$f(b^A b)$	b_1	b_2		$f(b)$	$f(b b^A)$	b_1	b_2
b_1	.5	b_1	.6	.4		b_1	.3	.2	
b_2	.5	b_2	.2	.8		b_2	.1	.4	
						$f(b^A)$.4	.6	
$f(b b^A)$	b_1	b_2							
b_1	.75	.33							
b_2	.25	.67							

Notice, however, that some uncertainty about the outcomes remains when the advice is perceived to be not perfectly accurate. This implies that the patient will not undoubtedly comply with the prescribed

treatment. It is still possible that he prefers to purchase (or obtain elsewhere) additional information (e.g. a second opinion).

(c) Perfectly accurate doctor

It is interesting to see that if the advice is perceived as perfectly accurate and there is some uncertainty in the consumer's a priori expectations, then the advice will always be followed.

	$f(b)$	$f(b^A b)$	b_1	b_2	$f(b)$	$f(b^A b)$	b_1	b_2	$f(b b^A)$	b_1	b_2
b_1	.9	b_1	1	0	b_1	.9	0	b_1	1	0	
b_2	.1	b_2	0	1	b_2	0	.1	b_2	0	1	
			$f(b^A)$.9	.1				

This is perhaps one of the most common situations, not because physicians are always so accurate but because the belief that 'the doctor knows best' is widespread and professional uncertainty often underestimated (see e.g. Wennberg et al., 1982).

(d) Completely inaccurate doctor

When the advice is perceived to be completely random, it will not affect the patient's subjective estimate of treatment outcomes. In such cases, he will not even seek (costly) advice.⁶

	$f(b)$	$f(b^A b)$	b_1	b_2	$f(b b^A)$	b_1	b_2
b_1	.7	b_1	.5	.5	b_1	.7	.7
b_2	.3	b_2	.5	.5	b_2	.3	.3

The examples given above serve to illustrate the fact that in this Bayesian framework consumer information has two distinct components:

1. prior information about effectiveness of medical care
2. information about the accuracy of medical advice

This second factor turns out to be at least as important in the demand for and the effect of information. It is worthwhile to explore this issue further in the case of medical care because the asymmetry in the first type of information between patient/consumer and doctor/agent will always remain and may even increase in the future with growing specialization. It may, however, be easier for consumers to obtain, or for regulators to provide, information of the second type. If indeed, as Pauly argues, doctors are able to generate demand for their own services through manipulation of the accuracy level of the information they provide, then it is clear that an important limiting factor on this ability is the accuracy perception of their clients. Consequently, the rather irreducible uncertainty about therapeutic performance of medical care may become less relevant to a discussion on the desirability and feasibility of competition in medical care if the consumers are reasonably well-informed about the degree of ignorance of the medical experts. Pauly (1978) puts it like this:

"The ironic conclusion is that one of the most useful types (and probably one of the least expensive types) of information that could be provided to patients is information on what is not known by medical

science and physicians." (p. 18)

About the relationship between these two types of information little can be conjectured a priori on a theoretical basis except that people without much knowledge about and experience with medical care will in general find accuracy of information hard to judge. There is, however, a tendency in the literature to observe that the typical medical lay people systematically underestimate professional uncertainty. Compare Breyer (1982): 'Casual observation would tend to suggest that many patients overestimate what medicine can do for them, thereby biasing the expected benefits of diagnosis upward.' (p. 155)

In the empirical analysis we will return to this point to see whether medical knowledge and perceived accuracy of medical advice are indeed negatively associated.

3.5. Conclusions and assessment of the theoretical model

Before we proceed with the empirical specification of this model, it is useful to summarize its main advantages and limitations compared to other economic models of the demand for health and medical care. The main distinguishing feature of this model is that it explicitly takes account of the agency relation which exists between the doctor and the patient by introducing a central role for information. Although individuals remain the final decision makers about medical care utilization, they are assumed to do this on the basis of their own beliefs and medical advice. Bayesian decision analysis combines these two into a posterior belief about the effectiveness of medical care. In this way, consumer sovereignty is no longer an on/off characteristic but varies continuously with the degree to which the prior expectations are altered into posterior beliefs. Secondly, if the consumer's medical information is assumed to be a positive function of the level of education, the Bayesian approach provides a more explicit mechanism to explain the positive effect of schooling on health and the negative impact on the demand for care. In Grossman's model, education, as a measure of general human capital, directly enters the individual's health production function as an efficiency-enhancing factor.

Finally, the analysis predicts that 'perceived accuracy of physician's advice' is an important determinant of the decision whether or not to consult a doctor. In this way, an attitude or belief variable about health and medical care, which is often found among the predisposing factors in behavioural models of health care utilization (e.g. Andersen, et al., 1975), is here explicitly derived from an economic model of utility maximizing behaviour.

Notes

1. This assumption is exactly opposite to the one made by e.g. Breyer (1982) in his analysis of rational purchase of medical care when H_0 is uncertain but b is known.
2. A normal distribution is chosen because it has some convenient properties in Bayesian preposterior analysis (see e.g. Hey, 1979).
3. Kihlstrom (1974a) actually assumes θ to vary continuously in order to make functions differentiable with respect to θ (see p. 103).
4. To avoid confusion, we do not adopt Kihlstrom's (1974a) terminology here. He calls expression (3.6) the expected utility of perfect information. For the sake of clarity, it is better to call it the expected utility with or under perfect information and to use the former term for expression (3.7) which indicates the expected utility difference between no (or only prior) and perfect information (see e.g. Hey, 1979).
5. It is also assumed that not b , but rather $\log b$ is normally distributed. The definition $b = e^{\epsilon}$, with ϵ normally distributed, is adopted to avoid negative values for b .
6. Interesting enough, a similar result was derived by Eeckhoudt, et al. (1985) who studied the effects of risk aversion on the demand for diagnostic test information by physicians.
7. Note the similarity between this interpretation and the hedonic approach adopted by Goldman & Grossman (1978) to distinguish between the quality and quantity of physician visits. They also define the total price of a physician visit (p) as the product of quality (q) and the quality-adjusted price (\hat{p}). However, in the hedonic approach, quality is not uncertain. It is traded-off against the quantity of visits in the production of health and consumers can substitute both inputs freely to maximize utility. Hedonic price theory is more suitable in order to explain the choice of physician when physician characteristics are available to measure his quality.
8. It is important to mention here what Pauly has termed a 'censoring problem'. More informed consumers will in general be less responsive to additional information, but they may in the first place also be less inclined to buy information and thereby be 'censored out' initially.

Chapter 4. Empirical analysis

"Potential observability seems to me to be an unnecessarily severe requirement. Hypothetical constructs such as 'permanent income' or 'human capital' are useful even if one cannot think of any procedure which would measure them directly" (Griliches, 1977, p. 9)

4.1. Description of data and Belgian health care characteristics

Because of the central role of consumer information in our theoretical model, it was necessary to look for a dataset which offered sufficient perspective for exploring this issue. The Health Interview Survey of the Belgian National Research Project on Primary Health Care conducted in 1976 among a representative sample of 1745 adults drawn from the Dutch-speaking (Flemish) population, showed several advantages in this respect. The stratified cluster sampling design and observation methodology are described in Foets et al. (1985) and Nuyens (1980).

Health care utilization and health and illness behaviour of the respondents during a total period of 9 weeks were registered in three health interviews about a 3 week recall period each. For the last two interviews, respondents were also asked to keep a health diary. Because the research project was largely set up from a sociological point of view, the questionnaires were designed to measure, apart from 'need' and 'enabling' factors, also so-called predisposing factors like attitudes, beliefs and knowledge which form the 'health perspective' of the individual (see e.g. Kohn and White, 1976). It is particularly this category of factors which is often included vaguely in the 'taste' variables in economic models, but becomes more explicit in the Bayesian approach to modelling the demand for information.

From secondary data sources, regional characteristics such as health care supply densities were added to the survey material. The stratified sampling procedure guaranteed that sufficient variation was

present according to degree of urbanization and health care supply density. The variables that were used for our analysis are described in appendix 4.A and will be discussed in more detail in the following sections. After listwise deletion of cases with missing values on these variables, the sample size retained for further analysis was reduced to $n = 949$. This reduction was primarily caused by the high non-response rate (19%) on the income variable. We did neither test nor correct for possible sample selection bias as a result of this non-response.

Due to the limited sample size, we have restricted our analysis to the two most frequently used types of medical care: about two-thirds of the individuals in the sample (63%) have used medicines during the observation period, and about one-third (36.1%) has consulted a general practitioner (Foets, et al., 1985, p. 184). We will, however, distinguish between medical consumption at the patient's own initiative and physician-prescribed utilization.

In Belgium, a public health insurance scheme covers medical expenses of virtually all inhabitants. In 1976, the cost-sharing for general practitioners and prescription drugs only consisted of a small fixed deductible amount (40 BF) per visit or per standard package delivered. A subgroup of high-risk beneficiaries, the widows, orphans, pensioners and invalids (the WOPI's) were exempted from this co-payment. Financial accessibility of health care therefore did not seem to be a large problem in 1976.¹ The same applies to geographical accessibility in a country with one of the highest general practitioner and pharmacy densities in the world (see OECD, 1985). The average level of GP contacts and in particular drug consumption per capita is quite high.

Before we estimate demand functions for information (i.e. GP consultations) and medical care (i.e. prescribed and non-prescribed medicines) in a structural equation model, we will first pay special attention to the measurement problems associated with two important determinants: health and knowledge, because: "the stock of health, like the stock of knowledge, is a theoretical concept; one that is difficult to define and quantify empirically" (Grossman, 1975, p. 165). The health and the information measurement 'modules' will be integrated in the full model in section 4.4.

4.2. Measurement of health

4.2.1. Methodology: the LISREL model

Obviously, the health measurement problem has received a great deal of attention among health economists and health services researchers in general. Almost without exception, health care utilization studies have found health status - or the variables used to measure it - to be

the single most important predictor. Moreover, Manning et al. (1982) have shown that more comprehensive health measurement leads to a substantial gain in explanatory power and better (unbiased) estimates of the impact of other determinants of the demand for medical care services.

One approach is to use more indicators of the several dimensions (physical, mental, social) of health in equations explaining utilization (see e.g. Brook et al., 1979, for an overview of health measures used in the Rand study). Another approach is to try to combine several variables into one 'health index' using factor or principal component analysis (see e.g. Levine and Yett, 1973). The development of accessible computer programmes has stimulated the use of latent variables in structural equation models. The convenient property of the so-called MIMIC methodology that latent variables can be characterized by both observable multiple indicators and multiple causes has not been left unnoticed by health economists. Robinson and Ferrara's (1977) example has stimulated a large number of applications of the latent variable technique, mostly in studies of the demand for health or health care. The list includes Van de Ven and Van der Gaag (1982), Wolfe and Van der Gaag (1981), Lee (1982), Hooymans and Van de Ven (1982), Van der Gaag and Wolfe (1983), Leu and Doppman (1984), Wolfe and Behrman (1984), Wagstaff (1986), Van Vliet and Van Praag (1987). In most of these studies one (MIMIC) health index is constructed, although sometimes this involves a two-step procedure. First a large number of indicators is combined into a smaller number of health status dimensions using factor or principal component analysis. These constructs are then used as indicators for the health index.

In the construction of these MIMIC health indices for individuals all too often the important distinction between the stock and flow component introduced by Grossman (1972) through his human capital framework tends to be neglected. The unobservability of the health capital stock as opposed to the flow of sick time in a certain period led researchers to assume that the latent variable represents the stock of health and is indicated by sick time and medical consumption. A similar distinction is made by Van de Ven and Van der Gaag (1982) but for a different reason. In adopting the Andersen et al. (1975) behavioural model of health care utilization, they considered 'permanent' health as an (unobservable) predisposing variable which influences the attitude towards health care utilization but is in itself no reason for seeking care. 'Number of days ill' reported by the respondent was used both as an indicator of permanent health and as a variable reflecting 'medical need' in the consumption equations.

Although our theoretical model does not make the distinction between a stock of health and a flow of sick time, we will distinguish between these two concepts in the health measurement module. In particular, we will hypothesize that not (permanent) health status but rather transitory distortions or illness periods throughout the observation

period determine the demand for information and care. The probability of occurrence of these distortions is, however, assumed to be increasing with reductions in the initial level of health H_0 . H_0 therefore only affects demand through its influence on these temporary health deviations. Our demand equations can then be interpreted as conditional upon the actual state of health of the individual at the beginning of the period (H_0) rather than as structural demand for medical care functions in which the desired health level (\hat{H}) should have a positive effect due to the derived demand hypothesis (see Wagstaff, 1986, p. 201).

The Primary Health Care Survey (from now on abbreviated as PHCS) included the following indicators of ill health (see Appendix 4.A for a more detailed description of the variables):

1. Health complaints (COMPL) reported during the 9 weeks period and classified into 14 general categories of perceived morbidity (see Nuyens, 1980, p. 80).
2. Number of restricted activity days (RAD) in the observation period.
3. Number of non-activity or in bed days (NAD) in the observation period.
4. A score on an index of behavioural impact of illness (DISAB) measured by questions about the ability to perform activities of daily living and feelings of well-being.
5. Number of serious chronic conditions and/or handicaps which the respondent was suffering from at the time of interview (CCOND).
6. Average frequency of GP visits in the past (PGPC).

In order to investigate whether these items represent observable indicators of one or more components of the (not directly observable) individual health status, we estimated a system of structural equations which is a special case of the general LISREL (Linear Structural Relations) model described in Jöreskog and Sörbom (1979). They developed the LISREL computer program for specifying, estimating and testing systems of linear structural relations. Their most general model consists of a structural part, a measurement part and a set of assumptions.

We will not first present the full model configuration but rather describe the relevant components as we go along. Because at this stage we only want to use the measurement part of the model, the structural part of the LISREL model (see equation 4.7) reduces to

$$\eta = \xi \quad \text{with} \quad \xi \sim N(0, \Psi) \quad (4.1)$$

where $\eta = a(m \times 1)$ vector of endogenous (latent) variables

$\xi = a(m \times 1)$ vector of disturbance terms

$\Psi =$ the $(m \times m)$ covariance matrix of ξ

In addition, we can write the following general measurement model for the η 's:

$$y = \Lambda \eta + \epsilon \quad \text{with} \quad \epsilon \sim N(0, \theta_\epsilon) \quad (4.2)$$

where η = a (mx1) vector of latent unobservable variables (e.g. health)

Λ = a (pxm) matrix of λ coefficients

y = a (px1) vector of observable indicators

e = a (px1) vector of errors of measurement

θ_e = a (pxp) covariance matrix of e

This LISREL measurement model can be seen alternatively as a confirmatory factor analysis model in which multiple indicators of one unobserved variable are used in order to correct for measurement errors (Jöreskog and Sörbom, 1979). The usual assumptions are that $E(y) = E(\xi) = E(\eta) = E(e) = 0$. Furthermore, the e 's (errors-in-variables) are assumed to be uncorrelated with the η variables ($E(\eta e') = 0$) and with the ξ 's (errors-in-equations) ($E(\xi e') = 0$). Maximum likelihood (ML) estimates of the λ -coefficients, their standard errors and the variances of the disturbance terms of θ_e can be obtained if the model is identified. The ML estimation method also assumes that the observed variables have a multivariate normal distribution.

Equation (4.1) assumes that the latent variable (or 'true score') is proportional to each of the indicators (or 'observed scores'), except for a random disturbance term (to allow for measurement error). The reliability of the indicators can be checked by means of the standardized solution measurement model. If we rescale appropriately, every relationship in (4.2) can be rewritten as

$$y_1 = \eta + e_1 \quad (i = 1, \dots, p) \quad (4.3)$$

$$\text{Dividing both sides by } \sigma_{y1} \text{ gives: } \frac{y_1}{\sigma_{y1}} = \frac{\eta}{\sigma_{y1}} + \frac{e_1}{\sigma_{y1}} \quad (4.4)$$

If we define standardized variables as $y^* = \frac{y_1}{\sigma_{y1}}$, we get

$$y^* = \frac{\sigma_{\eta}}{\sigma_{y1}} \cdot \frac{\eta}{\sigma_{\eta}} + \frac{e}{\sigma_{y1}} \quad (4.5)$$

$$\text{or } y^*_1 = \lambda^*_1 \eta^* + \frac{e_1}{\sigma_{y1}} \quad (4.6)$$

$$\text{where } \lambda^*_1 = \frac{\sigma_{\eta}}{\sigma_{y1}}$$

The reliability coefficient ρ of each indicator is now defined as the standardized λ^*_1 -coefficient squared. It gives the ratio of the true score variance and the observed score variance. Tests of hypotheses about the reliability can only be performed when more than two indicators are available and overidentifying restrictions can be imposed on the model structure. Examples will be provided in the following sections.

Furthermore, the standardized solution can be used to test whether the indicators are congeneric (i.e. measure the same theoretical variable), tau-equivalent (i.e. have the same true score but different error variances) or parallel (i.e. have identical true scores and

error variances) (for an example of psychometric testing with LISREL, see e.g. Jöreskog and Sörbom, 1979, Chapter 3).

4.2.2. A latent health vector

4.2.2.1. Model specification and testing

In this section, we will proceed stepwise to the construction of a LISREL health measurement model for our dataset. First, we will compare the goodness-of-fit and reliability performance of a two-component model with the usually estimated one-component model. Secondly, by adding health-determining variables, we can check the validity of the constructs as measures of individual health in section 4.2.3. The correlation matrix of the variables used in this section is given in appendix B, table B.1. Some of the variables (i.e. RAD, NAD, DISAB, PGPC, IHIS) have been transformed by taking the natural logarithm of the variable plus one. This (arbitrary) transformation procedure was applied to reduce the skewness of the distribution of these variables and the impact of some outlier values.

In order to make such a system of equations like (4.2) estimable, the problem of the scale of the latent variables has to be solved. This is in fact an identification problem because multiplying each η -variable by some constant and dividing the λ 's by the same constant produces the same y -variables. As such, an infinite number of different solutions to this system can be produced. Two normalization procedures exist to remedy this scale problem (see Jöreskog and Sörbom, 1979). One is to fix one λ -coefficient for each latent variable equal to a constant. For instance, by fixing one λ equal to 1, the latent variable obtains the same unit of measurement as one of the indicators. The alternative identification procedure is to fix the variances of the unobservable variables in this particular case (the diagonal elements of ψ) equal to one. An advantage of the second method is that it directly gives the standardized solution of the model which facilitates testing of psychometric properties (see section 4.2.1).

We first tested the hypothesis that all six variables listed in the previous section can be seen as indicators of one underlying latent variable health. The results are presented in table 4.1, Model 1A ($p = 6$, $m = 1$). Table 4.1 contains both unstandardized estimates of the λ and ψ coefficients (first column) and standardized results (second column) for two model versions (1A and 1B). In model 1A all six variables are assumed to be indicators of one unobservable variable 'good health' because the coefficient of COOND is set equal to -1. All coefficients have the expected sign and small standard errors but the maximum likelihood parameter estimates and the t -values have to be interpreted with caution due to the non-normal distribution of most y -variables.

Table 4.1. Estimation results of health measurement models (6 indicators)

y	MODEL 1A (one health variable)		MODEL 1B (two health variables)	
	$\hat{\lambda}$	(1) $\hat{\rho}$ (2)	$\hat{\lambda}$	(1) $\hat{\rho}$ (2)
CCOND	-1.000	.279	-1.000	.269
PGPC	-1.054 (9.6)	.308	-1.080 (10.5)	.314
COMPL	-1.238 (10.4)	.339	-1.194 (10.9)	.384
LRAD	-1.165 (10.1)	.231	-.590 (12.5)	.256
LNAD	-1.236 (10.4)	.237	-.650 (13.4)	.310
LDISAB	-1.748 (11.6)	.488	-1.000	.735
ψ_{11}	.199 (6.3)	1.00	.269 (6.9)	1.00
ψ_{22}	-	-	.735 (12.1)	1.00
ψ_{21}	-	-	.326 (10.4)	.733
χ^2		84.6		28.6
df		9		8
AGFI		.924		.971

Notes:

- 1) Approximate absolute t-values ($\hat{\lambda}$ coefficient divided by standard error) are given in parentheses. When no t-value is reported, this means that the parameter was fixed to be equal to the given value.
- 2) The reliability coefficients were calculated from the estimated $\hat{\theta}_e$ variances using the following formula:

$$\phi = 1 - \hat{\theta}_e$$

Boomsma (1983) has examined the robustness of LISREL against non-normality by means of a Monte Carlo study. He found that, for a sample size of 400, LISREL is rather robust against categorization but not against skewness. Large skewnesses affect the confidence of parameter estimates and the χ^2 goodness-of-fit statistic. Overestimation of standard errors of parameters causes confidence intervals to be too large and therefore tests based on these intervals too conservative. On the other hand, skewness increases the chi-square test statistic and therefore too often leads to model rejection. These consequences will have to be kept in mind because skewness is an important characteristic of the dependent variables in our sample (see appendix 4.A). The suggestion of Muth  n (1983) and others to use polychoric and polyserial rather than the ordinary product-moment correlations as input measures of association for LISREL estimation does not seem to be a good alternative in our case because it is based on the assumption that the underlying latent variables have a normal distribution. This assumption seems to be very questionable in the case of medical consumption and health status which are intrinsically skewedly distributed across individuals.

The reliability coefficients given in the second column ($\rho = \lambda^2 = 1 - \theta_{\epsilon}$) are not very high. Moreover, the χ^2 test statistic rejects the null hypothesis that the observed correlations between the variables could have been generated by the specified causal model. However, this test statistic is proportional to the sample size and therefore very often leads to model rejection in the case of a large sample size like ours ($n = 949$). Jöreskog and Sörbom (1981) warn that: "large sample sizes and departures from normality tend to increase χ^2 over and above what can be expected due to specification error in the model." (p. 1.39) It is therefore recommended to use the χ^2 measure in such cases rather for nested model comparison than for the evaluation of just one model (Jöreskog, 1971).²

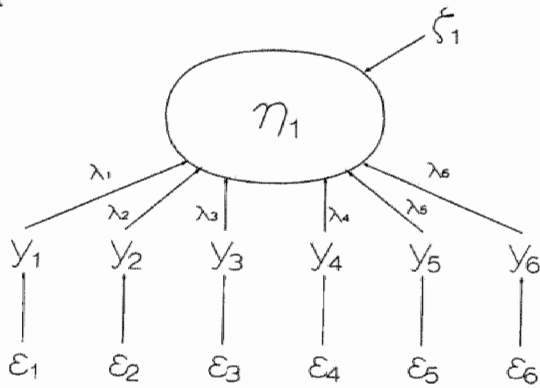
The goodness-of-fit index, adjusted for the degrees of freedom (AGFI), measures the extent to which the model accounts for the variances and covariances between the observable variables in the model ($0 \leq \text{AGFI} \leq 1$). Its advantage is that it is independent of the sample size and reasonably robust against departures from normality, but on the other hand its distribution is unknown. It can therefore also be used only for model comparisons (see Jöreskog and Sörbom, 1981).

In order to test our proposition that health consists of two distinct parts, i.e. a fairly permanent health state and the transitory health distortions during a certain period, we explored the consequences of subdividing the health indicator variables between these two latent variables (Model 1B, $p = 6$, $m = 2$, see also Figure 4.1). Clearly, the presence of chronic illnesses or impairments (CCOND) and the average visit rate of the GP in the past (PGPC) can be regarded as indicators of the general level of health. On the other hand, restricted (LRAD) and non-activity days due to illness in the 9 week period and the behavioral impact index (LDISAB) which measures sleeping problems, worries and normal activity interruption in the same period are indicators of the amount and intensity of sick time. They were therefore expected to load on the transitory health component. The only interpretation problem was caused by the variable "number of different health complaints expressed by the respondent during the 9 weeks" (COMPL). Although we conjectured it to be a measure of the flow of sick time, it turned out to be a better indicator of permanent (η_1) than of transitory health (η_2).

Model 1B in table 4.1 shows the results when CCOND, PGPC and COMPL are taken to be indicators of permanent health (η_1) and when LRAD, LDISAB and LDISAB indicate transitory health (η_2). As can be seen from the reliability coefficients and χ^2 test statistic, this model fit represents a substantial improvement to model 1A. The statistical significance of the difference between the Chi-squared tests was not formally tested because both models are not hierarchical (see Bentler and Bonett, 1980), and we did not perform a likelihood ratio test of the non-nested hypothesis that 1B has a better fit than 1A. But it is clear from a comparison of the goodness of fit statistics that 1B fits

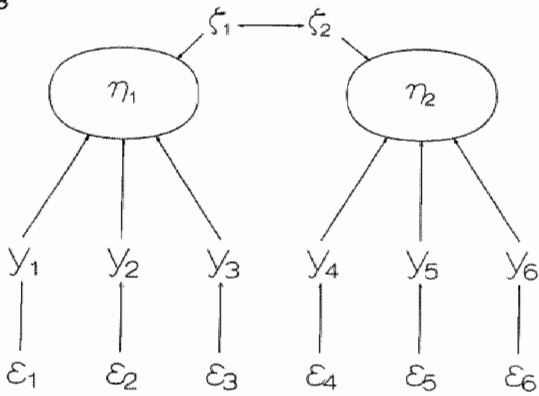
Figure 4.1. Path-diagram for the sequence of health measurement models

a. Model 1A



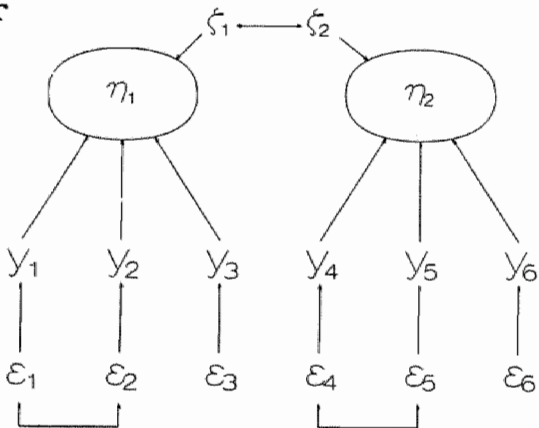
Restrictions
 $\lambda_1 = -1$

b. Model 1B



Restrictions
 $\lambda_1 = -1$
 $\lambda_6 = -1$

c. Model 1C



Restrictions
 $\lambda_1 = -1$
 $\lambda_6 = -1$

d. Model 1D = Model 1C but with additional restrictions $\lambda_3 = \lambda_4$ and $\theta_{\epsilon 5} = \theta_{\epsilon 4}$

the data better than 1A.

Because more information was available on the (total) complaints variable (COMPL), we attempted to discriminate between the more chronic and acute complaints. Respondents had been asked how long they had been bothered by each of these complaints. Because the chronicity of complaints varied considerably between the fourteen categories of perceived morbidity, we divided COMPL into three subcategories: Number of complaints in any of the 'acute' categories (ACUTE = less than 20% is reported as chronic, i.e. lasting for over a year already), in the 'chronic' disease category (CHRONIC = more than 45% chronic) or in the intermediate range (ACCHR = between 20 and 45% chronic) (see appendix 4.A for a more detailed description).

Table 4.2. Estimation results of health measurement models (8 indicators)

	MODEL 2A		MODEL 2B	
	$\hat{\lambda}$	$\hat{\rho}$ (2)	$\hat{\lambda}$	$\hat{\rho}$ (2)
CCOND	-1.000	.177	-1.000	.208
PGPC	-1.088 (8.9)	.187	-1.017 (9.2)	.215
CHRON	-1.123 (9.4)	.224	-1.144 (9.8)	.272
ACCHR	-1.312 (10.1)	.305	-1.310 (10.4)	.357
ACUTE	- .604 (6.2)	.065	- .297 (6.6)	.060
LRAD	-1.251 (9.9)	.278	- .634 (13.5)	.273
LNAD	-1.318 (10.1)	.308	- .688 (14.4)	.322
LDISAB	-1.836 (11.2)	.597	-1.000	.679
ψ_{11}	.177 (6.0)	1.00	.208 (6.2)	1.00
ψ_{22}	-	-	.679 (12.3)	1.00
ψ_{21}	-	-	.310 (10.1)	.824
χ^2	138.2		108.6	
df	20		19	
AGFI	.934		.942	

Notes: see table 4.1

Table 4.2 reports the results of the estimations with this extended set of indicators. The attempt to distinguish better between permanent and transitory health did not prove very successful. Model 2A is the one-health-component analog of 1A and 2B the two-component model with only ACUTE as an indicator of transitory health. However, ACUTE has a very low reliability coefficient. The lower estimated reliabilities of the other indicators and the worse goodness-of-fit statistics do not give rise to any support to the hypothesis that this solution improves the discrimination between two health components.

Moreover, the standardized covariance between the two latent variables (ψ^*_{21}) in model 2B is .824 and was only .733 in model 1B. When this correlation is equal to one, both constructs can be seen as representing one underlying factor. This hypothesis can be tested formally by imposing the restriction that $\psi^*_{21} = 1$ and re-estimating the model. The difference in χ^2 statistics between the restricted and unrestricted model then itself has a χ^2 distribution with one degree of freedom (Sarlis and Stronkhorst, 1984, p. 276). When we performed this 'parameter nesting' to model 1B, the more restricted model yielded a χ^2 statistic of 84.6. Compared to the 28.6 χ^2 value of model 1B, the increase (56) leads us to reject the hypothesis that both health components actually measure the same thing. We, therefore, preserved the distinction between permanent and transitory health and model 1B was selected as our preferred model specification for further development.

4.2.2.2. Model modification

As it stands, however, the χ^2 test still rejects the hypothesis that this model has generated the observed correlation matrix. The LISREL V program does produce also so-called 'modification indices' for each constrained parameter. These indices are based on the first and second derivatives of the likelihood function with respect to the zero constrained parameters. They equal the expected minimal decrease in the value of the model's χ^2 statistic if the zero restrictions on these parameters are relaxed. As such they provide an indication as to the possible source and location of misspecifications in the model. Clearly, modification indices are a purely statistical criterion to be used cautiously and in combination with theoretical considerations.

Table 4.3. Modification indices for constrained parameters in model 1B

	LAMBDA (λ) MATRIX		THETA EPSILON (θ_e) MATRIX				
	PH η_1	TH η_2	CCOND γ_1	PGPC γ_2	COMPL γ_3	LRAD γ_4	LNAD γ_5
CCOND	-	.219	-				
PGPC	-	.539	8.180	-			
COMPL	-	1.052	2.536	0.704	-		
LRAD	1.472	-	5.519	0.081	2.037	-	
LNAD	0.000	-	0.006	0.187	0.087	6.261	-
LDISAB	.363	-	3.770	1.859	1.458	0.000	1.423

By way of example, we have reproduced the modification indices of model 1B in table 4.3. Because there are six λ parameters constrained to zero in this model plus all the non-diagonal elements of θ_e , two

types of model modifications based on these indices can be contemplated: indicators loading on both latent variables or correlations between the disturbances of the measurement equations. The first possibility does not seem to cause any problems in this case, but the measurement errors of two indicators seem to be correlated for both latent variables: PGPC and CCOND for η_1 , and LRAD and LNAD for η_2 . That the latter two variables produce correlated errors in measuring transitory health seems logical because they both have (the logarithm of) a number of days as the unit of measurement. When the zero covariance restriction on their disturbance terms was relaxed, the χ^2 dropped to 19. Despite the significant fit improvement, it still means rejection at the 5% level of significance (with 7 degrees of freedom).

It was only when, in addition, in model 1C (table 4.4 and figure 4.2) the error terms of CCOND and PGPC (besides those of LRAD and LNAD) were allowed to covary in model 1C (table 4.4) that a $\chi^2_{(6)} = 7.9$ was obtained which did not exceed the 5% critical value of 12.59. Apparently, the presence of one or more chronic conditions or handicaps and the average use of GP services show some correlation which is insufficiently explained by one common component ('permanent health') with the number of complaints. Health model 1C in table 4.4 is a specification which, according to the goodness-of-fit tests, fits to our data very well, i.e. the predicted correlations between the observed variables deviate very little from the actual correlations.

Table 4.4. Estimation results of health measurement models 1C and 1D

	MODEL 1C		MODEL 1D	
	$\hat{\lambda}$	(1) $\hat{\rho}$ (2)	$\hat{\lambda}$	(1) $\hat{\rho}$ (2)
CCOND	-1.00	.221	-1.00	.222
PGPC	-1.087 (10.2)	.261	-1.084 (10.3)	.261
COMPL	-1.353 (9.8)	.405	-1.351 (9.8)	.405
LRAD	-.653 (12.2)	.289	-.684 (15.0)	.316
LNAD	-.709 (13.0)	.341	-.684 (15.0)	.316
LDISAB	-1.000	.678	-1.00	.674
ψ_{11}	.221 (5.9)	1.00	.222 (5.9)	1.00
ψ_{22}	.678 (11.5)	1.00	.674 (11.5)	1.00
ψ_{21}	.301 (9.5)	.778	.302 (9.5)	.780
θ_{e54}	-.086 (3.0)	-.086	-.088 (3.0)	-.088
θ_{e21}	.108 (3.4)	.108	.108 (3.4)	.108
χ^2		7.94		9.43
df		6		8
AGFI		.989		.990

Notes: see table 4.1

There is one last hypothesis about this model which can be tested by the LISREL framework using the same straightforward nesting procedure. We would expect one day of non-activity to be worse in terms of ill (transitory) health than one day of restricted activity as reported by the individuals. In other words, the same change in the 'true' health score would imply a higher number of restricted than non-activity days. Consequently, the λ -coefficient for LRAD would have to be larger than that for LNAD. But the results of model 1C indicate exactly the opposite situation ($|\lambda_{42}| < |\lambda_{52}|$).

We can state the null hypothesis more formally as: $\lambda_{42} = \lambda_{52}$. The equality of these two parameters can be tested using the same difference (in χ^2) test statistic as before. The difference test statistic was found to be equal to

$$D = \chi^2(7) - \chi^2(6) = 8.88 - 7.94 = 0.94$$

which means that imposing the equality restriction does not raise the χ^2 statistic significantly and that the hypothesis cannot be rejected. It implies that in terms of classical test theory both indicators can be seen as congeneric instruments (see section 4.2.1).

Moreover, when the additional restriction was imposed that both indicators have also equal variances, i.e. $H_0 : \theta_{\epsilon 44} = \theta_{\epsilon 55}$, we obtained model 1D presented in table 4.4 and described in figure 4.1. It can be seen that the test statistic now equals

$$D = \chi^2(8) - \chi^2(7) = 9.43 - 8.88 = 0.55$$

which again implies that also this restriction cannot be rejected, and we have to conclude that LRAD and LNAD are even parallel instruments of transitory health. In other words, in our sample the number of restricted and non-activity days reported by the respondent not only measures the same thing but they also appear to represent measurements on the same scale (see Jöreskog, 1979). Adding these restrictions to the measurement model for health leaves the estimates of the other coefficients virtually unaffected. Model 1D will be retained for further analysis because it is the most restricted measurement model which fits the data and conforms to our theoretical distinction between two health components.

4.2.3. A MIMIC health model

In this section we extend the basic measurement model 1D with a structural equation part. By specifying a model of causal relationships between the latent variables and their determining factors, a Multiple Indicator Multiple Causes model is obtained (see Jöreskog and Goldberger, 1975). Although we did not derive a demand for health function from the theoretical model, and initial permanent health H_0 is assumed to be predetermined, we will estimate a sort of auxiliary equation to relate permanent and transitory health to a number of

background characteristics in order to investigate their validity.

We specify, in addition to (4.2) and instead of (4.1) the following equation:

$$\eta = B\eta + \Gamma x + \xi \quad (4.7)$$

with η = a (2x1) vector of latent health variables = [PH, TH]'

x = a (5x1) vector of exogenous variables

= (AGE, FEM, EDUC, PSPR, LIHIST)'

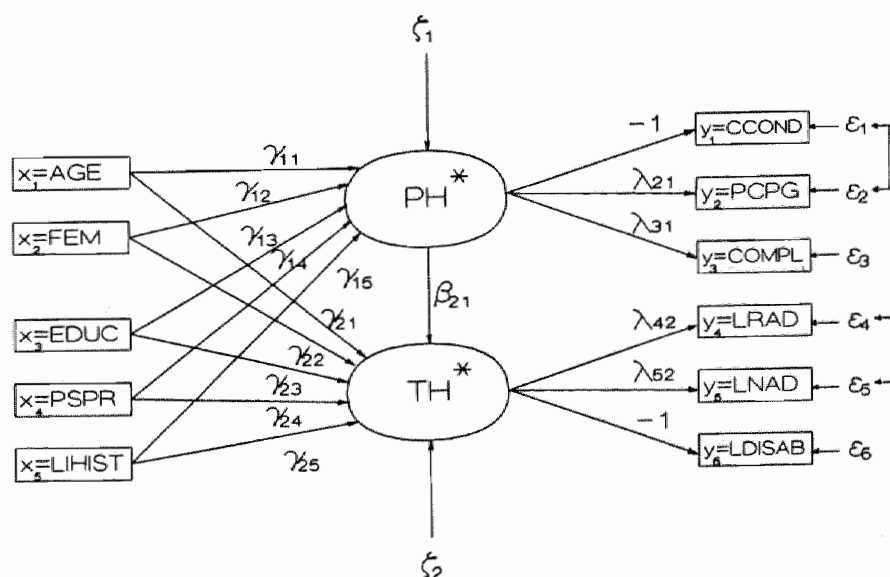
B = a (2x2) matrix of beta-coefficients (with zero diagonal elements)

Γ = a (2x5) matrix of gamma-coefficients

ξ = a (2x1) random disturbance vector ($\xi \sim N(0, \Psi)$)

We expected health to be negatively related to age, female gender (FEM), the number of psycho-social problems (PSPR) (at home, on the job or in relations with friends and relatives) reported by the respondent and to his or her illness history (LIHIST, see appendix A), and positively related to the level of education. With respect to the relation between the η 's, it is assumed that permanent health (PH) has a direct positive effect (β_{21}) on transitory health (TH). In a path-diagram, the specified MIMIC-model can then be depicted as in Figure 4.2.

Figure 4.2. Path-diagram of MIMIC-health model



The maximum likelihood estimation results for this specification are given in table 4.5. All sign expectations for the effects of the exogenous determinants on PH are confirmed except for EDUC which has a non-significant coefficient. PH also shows a large positive impact on TH (β_{21}).

The findings with respect to the level of transitory health are more peculiar. Except for PSPR, the coefficients of all other determinants show the opposite sign from that in the PH-equation! The effects of EDUC, PSPR and LIHIST are small and only significant at the 10% level. But the results clearly show that, given their generally inferior permanent health status, women and older persons are less likely to report sick time, *ceteris paribus*. Because the latter is primarily measured as interruption of normal activities, this may have something to do with the kind of activities performed by these groups of people. If the major part consists of household work "which has to be done some way or another even if one is not feeling well", this may contribute to an explanation of this remarkable result.

However, these direct effects (γ_{21} , for $i = 1, \dots, 5$) are in general much smaller than the indirect effects through PH on TH ($\beta_{21} * \gamma_{1i}$, for $i = 1, \dots, 5$) which implies that the total effects (direct plus indirect) often are of the opposite sign. This decomposition into direct and indirect effects again brings out the importance of distinguishing between the permanent stock and transitory flow component of health: not only the indicators but also the 'causes' of health seem to indicate this distinction.

Table 4.5. Estimation results of MIMIC health model of figure 4.2

MEASUREMENT EQUATIONS		STRUCTURAL EQUATIONS	
y	$\hat{\lambda}$	PH*	TH*
OCOND	-1.0		
PGPC	-.997 (11.9)	AGE -.195 (7.5)	.182 (4.4)
COMPL	-1.057 (11.5)	FEM -.142 (6.2)	.106 (3.0)
LRAD	-.680 (15.5)	EDUC -.012 (.49)	.057 (1.73)
LNAD	-.680 (15.5)	PSPR -.110 (4.9)	-.061 (1.83)
LDISAB	-1.0	LIHIST -.099 (4.4)	.063 (1.90)
		PH*	-
			1.319 (8.8)
$\hat{\theta}_{e21}$.054 (1.73)	R ²	.279
$\hat{\theta}_{e54}$	-.086 (3.0)		.645
χ^2	= 171.4		
df	= 27		
AGFI	= .944		

Notes: see table 4.1

Finally, note that, despite the addition of the structural part of the model, the measurement model estimations are virtually unaffected. On the other hand, the extension did influence the goodness-of-fit of the model: the χ^2 test now rejects the hypothesis that this simple linear and additive model specification can account for the correlations between the observed variables. But even on the basis of these five general background characteristics the percentages of explained variance (28% for FH and 65% for TH) are relatively high already and this leads us to conclude that this MIMIC model is useful as the health measurement 'module' to be implemented in the total structural model in section 4.4.

4.3. Measurement of information

4.3.1. Previous studies

In contrast to the vast literature on health measurement in medical care demand models, the issue of how to measure consumer knowledge-ability about health and medical care matters has received relatively little attention from economists. In testing his information-manipulation theory, Pauly (1980) used 'education of the household head' as a proxy for the prior stock of consumer information and the 'degree of urbanization of their place of residence' as a measure of the difficulty that individuals have in determining the accuracy of the physician advice. He hypothesized that it is easier to obtain such information in non-metropolitan areas where there are relatively fewer physicians to choose from.

Such crude proxies are also used by Pauly and Satterthwaite (1981) to investigate the relationship between consumer information about primary care physician's reputations and their pricing behaviour. Regional informational levels and search possibilities of consumers are assumed to be negatively related to the total number of physicians within a given market area, the proportion of residents who have moved in that area the past five years and the proportion of households headed by females with less possibilities for social contact and communication. They admit themselves that: 'A primary data collection project that directly measures consumer information would have the potential for yielding more definite results' (p. 496).

A slightly less indirect approach towards measuring the medical knowledge and expertise of consumers has been to assume that members of a family with at least one member in a medical occupation are better informed. In contradiction with the demand inducement hypothesis, Hay and Leahy (1982) found these medically informed patients to have the same or even higher utilization rates for physician office visits than individuals without family member with some medical education. Using a similar approach, Bunker and Brown (1974) had shown earlier that

doctors and their spouses had either higher or the same utilization rates for surgical procedures as other professional groups when standardized by age and sex. The problem with both studies is that it is unclear to what extent the proxies used for consumer medical information do not also capture other factors than medical knowledge. Moreover, no distinction is made between the two components of information derived from the Bayesian analysis.

An attempt to measure the individual sophistication or knowledgeability about the medical care delivery system in a direct manner was made in the Health Insurance Study of the Rand Corporation. Newhouse et al. (1981) administered a 10-item questionnaire, covering a range of information about the medical care delivery system, to a sample of 4976 non-elderly persons. These items refer to whether consumers are knowledgeable both in choosing medical care providers and in making decisions at the time when services are used. By means of a factor analysis, eight of these items were selected to score a consumer sophistication summary scale by counting the number of correct responses. This scale correlated substantially with education ($r = .40$) and showed a small but significant positive association with the number of physician visits during the prior year. Age and female sex were positively and race negatively related to consumer sophistication. Correlations with health status measures were low and inconsistent. Perhaps the most important finding with respect to our own analysis was: "The consistent negative and significant correlations with the belief and attitude measures suggest that consumers who tend to be more knowledgeable regarding medical care services also tend to view doctors and their services less favourably" (Newhouse et al., 1981, p. 323). This evidence points in the direction of a negative relationship between a priori medical knowledge and perceived accuracy of medical advice. Finally, the authors rightly stress the difficulties in examining the relationship between knowledgeability and overall utilization: 'because a more sophisticated consumer could utilize more frequently when appropriate and less frequently when inappropriate' (p. 324). This indeterminacy is confirmed in another study of the Rand researchers (Manning et al., 1982). The consumer sophistication scale is indeed found to have an insignificant (negative) effect on annual outpatient expenditures for health care when a large number of other variables are controlled for in the regression equation.

4.3.2. A MIMIC model of consumer information

It seems logical to take the latter analysis one methodological step further by investigating whether the relevant consumer information concepts can adequately be modelled as latent variables in a structural equation model explaining health care utilization. While some authors have concentrated on proxy determinants of consumer information and others on proxy indicators to measure it directly, we will

combine both approaches while preserving the distinction between consumer knowledgeability of health and medical care, and perceived accuracy of general practitioner advice. The former is assumed to be inversely related to consumer prior uncertainty before seeking advice or therapy, and the latter represents the information on (or perception of) the accuracy of professional advice.

4.3.2.1. Medical knowledgeability

With regard to medical knowledge, the FHC survey included 43 test items. These questions could be subdivided into four broad categories: 14 items were related to purely theoretical medical knowledge which does not have much relevance for the individual behaviour (TEST 1); 16 items measured aspects of theoretical knowledge which potentially could have some practical relevance (TEST 2); 11 items referred to very practical knowledge about what action one should undertake in certain circumstances (e.g. first aid or preventive activities) (TEST 3); finally, 2 items asked for some knowledge about the organization of the medical care delivery system (whether doctors need a licence to practice medicine and whether it is possible to see a specialist without a referral from the GP) (TEST 4). These four types of knowledge were scaled by summing the scores on the individual items. An answer was coded 0 if false, 1 if the respondent answered 'do not know' or did not answer, and 2 if correct. The four knowledge scales are highly intercorrelated (see appendix 4.B).

A simple measurement model with all four TEST variables used as indicators of the same underlying latent variable was not rejected by the χ^2 test (see model 1 in table 4.6). TEST 4 has a low reliability ($\rho = .104$), mainly because this score shows less variability as it is based on only 2 items. In model 2 it was deleted but this did not have much impact on the performance of the other indicators. We therefore decided to retain TEST 4 on theoretical grounds and to select the model 1 specification as the measurement model for a priori consumer knowledgeability about health and medical care.

Although the latter concept is exogenously given in the theoretical model of Chapter 3 by the parameter ϕ of the prior distribution, we will attempt to explain it by means of a number of background characteristics. As such, the index becomes again determined by both multiple indicators and multiple causes. Simultaneously, the validity of the construct can be checked from the signs of the coefficients of the determining factors. We will regard knowledge as a human capital stock which is built up initially through formal education early in the lifecycle, accumulates with experience but depreciates with ageing. It is therefore expected to be positively related to the level of schooling (EDUC) and negatively to age (AGE). Investment in medical knowledge is assumed to take place through own experience and learning from the experience of others. To operationalize this learning from

Table 4.6. Estimation results of measurement models for medical knowledge

	MODEL 1		MODEL 2		MIMIC MODEL	
	$\hat{\lambda}$	$\hat{\rho}$	$\hat{\lambda}$	$\hat{\rho}$	$\hat{\lambda}$	$\hat{\rho}$
TEST 1	1.00	.542	1.00	.551	1.00	.567
TEST 2	.941 (15.5)	.480	.920 (14.9)	.466	.902 (17.2)	.461
TEST 3	.914 (15.3)	.453	.912 (14.9)	.458	.890 (17.1)	.449
TEST 4	.438 (8.4)	.104	-	-	.421 (8.6)	.101
χ^2	2.77		0			
df	2		0			
AGFI	.992		1.00			

Notes: see table 4.1

others through communication of experience, we use the size of the household (HHSZ), the number of direct family members with a medical or paramedical occupation (MEDFAM) and the socio-economic status as measured by the occupational level of the head of the household (SES). Finally, we control for possible sex differences by including a dummy for females (FEM). These are the x-variables in the following LISREL structural equation model:

$$\eta = \Gamma x + \xi \quad (4.8)$$

where η = a latent variable (knowledge)

x = a (6x1) vector of exogenous variables

= (EDUC, SES, MEDFAM, HHSZ, AGE, FEM)'

Γ = a (1x6) vector of coefficients

ξ = a random disturbance term ($\xi \sim N(0, \psi)$)

and η is measured by a model like equation (4.2) with $y' = [\text{TEST 1, TEST 2, TEST 3, TEST 4}]$. The estimation results of this model are presented in table 4.6 (measurement part) and in table 4.7 (structural part). From the estimated λ -coefficients it can be seen that the measurement model is hardly affected by the extension of the model with a structural part. The estimated γ -coefficients in table 4.7 show that education, socio-economic status and medical family members improve medical knowledge significantly while older people and males are less knowledgeable. Household size does not have a significant effect on knowledge production. Despite the rather high percentage of explained variance (38%) in knowledge and a very high AGFI (.985), the model is rejected by the χ^2 statistic.

In order to illustrate the differences and similarities between the latent variable model and the ordinary maximum likelihood regression model, we have presented the regression results in the second column of table 4.7. The dependent variable is TESTTOT which is the sum of

Table 4.7. Estimation results of the 'medical knowledge' equation

	MIMIC MEDKN	REGRESSION TEST TOT	REGRESSION UNCERTAIN	REGRESSION CORRECT
EDUC	.246 (7.2)	.278 (7.1)	-.190 (4.5)	.269 (6.9)
SES	.206 (6.5)	.236 (6.5)	-.188 (4.9)	.243 (6.8)
MEDFAM	.075 (3.0)	.089 (3.1)	-.077 (2.5)	.095 (3.4)
HHSZ	.011 (.42)	.016 (.56)	-.071 (2.3)	.047 (1.6)
AGE	-.064 (2.3)	-.076 (2.4)	.092 (2.7)	-.094 (3.0)
FEM	.062 (2.6)	.079 (2.9)	-.035 (1.2)	.067 (2.4)
R ²	.378	.282	.182	.298
x ²	45.9	0	0	0
df	20	0	0	0
AGFI	.985	1.0	1.0	1.0

Notes: see table 4.1

the 4 indicators used in the MIMIC model. The coefficients cannot directly be compared due to the different units of measurement but the t-values show that the results are almost identical. The only substantial difference is the lower percentage of explained variance in the total sum score (28%). This reflects the fact that, through 'purging' the observed variables from a certain amount of measurement error in the MIMIC model, the remaining variance in the 'true score' of knowledge is better explained by this set of exogenous variables. In this particular case, this seems to be the only gain from treating knowledge as a latent factor rather than as a simple sum score of tests.

Finally, we explored the consequences of our arbitrary choice of test scoring method. Recall that we constructed TESTTOT by summing the score 2 if the answer was correct and 1 when the answer was 'don't know'. One could argue that, in order to measure a priori uncertainty (the parameter ϕ), the 'don't know' answer should be treated in the same way as an incorrect answer or should even be treated separately. Perhaps no answer gives a better indication of uncertainty than a wrong answer. We therefore examined the performance of two alternative knowledge variables: CORRECT was defined as the sum of all correct answers on the 43 items, and UNCERTAIN was constructed as the sum of all 'don't know' answers. Correlations, range and mean values of these variables and TESTTOT are presented in table 4.8.

The correlations between these constructs are very high and this similarity is also confirmed by the regressions presented in the last two columns of table 4.6. UNCERTAIN is inversely related to the determinants of the medical knowledge scores and, with the exception of household size (HHSZ), all variables show similar relative effect magnitudes and signs in each of the three equations. We therefore

Table 4.8. Correlations, range and mean values of knowledge scores

	CORRELATIONS			RANGE	MEAN
	TEST TOT	CORRECT	UNCERTAIN		
TEST TOT	1.0			31-86	53
CORRECT	.89	1.0		0-43	19
UNCERTAIN	-.54	-.86	1.0	0-43	14

conclude that our knowledge measurement is fairly robust with respect to alternative scoring procedures and we will continue to work with the four-indicator and six-causes MIMIC model in what follows.

4.3.2.2. Perceived quality of GP's

The empirical operationalization of the second (Bayesian) component of information, the perceived accuracy of information provided by general practitioners, is somewhat more difficult because no examples are available in the health economics literature and the theory does not offer much empirical guidance. What we actually would like to measure is the analog of the likelihood function in Bayesian decision theory which represents the sample information. In our situation, this could be captured by probabilistic statements like: 'In x% of the cases, the diagnostic or therapeutic information given by the doctor turns out to be correct.'

A number of statements approximating this example were included in the FHC survey questionnaire. The respondents were asked whether they agreed or disagreed with 34 statements evaluating general practitioner services. Most of these were not used in the analysis because they referred to aspects such as accessibility or general satisfaction which were not relevant for our purposes.

Twelve items were selected which referred to an evaluation of either the diagnostic accuracy or the therapeutic effectiveness of general practitioners. A translation of these items is listed in appendix C. The answer categories to these statements range from 0 (= disagree) to 4 (= agree) which is an ordinal or ranking scale over which the variables are not normally but at least rather symmetrically distributed (see appendix C for frequency distributions). The correlation matrix of these items is given in table B3 in appendix 4B.

In order to investigate what proportion of the variation expressed by the scores on these 12 items could be reproduced by a smaller number of underlying factors, we performed a factor analysis using the (SPSS-X) method of principal components for factor extraction. The objective is to find linear combinations of the observed variables which account

for the largest proportion of sample variance. Two factors with an eigenvalue larger than one were obtained which, after varimax rotation, produced the factor loading matrix presented in table 4.9.

It can be seen that these two components account for almost half (46%) of the total variance expressed by all items. The items divide equally over the two factors with 6 items loading on each factor. With the exception of item EVGP7 which loads positively on both factors, the factors distinguish between positively and negatively formulated items. Therefore, we arrive at the remarkable conclusion that stating the proposition in a positive or in a negative way seems to make more difference than the content of the proposition! For example, EVGP3 and EVGP10 are in fact the positive and the negative formulation of the same statement ('GP's do or do not examine you thoroughly') but they do not load on the same factor. We could not interpret these two components otherwise than as the inverse concepts 'distrust about GP quality' and 'confidence in GP quality'. However, by construction (the rotation procedure) both principal components are orthogonal and cannot be used as (positive and negative) indicators of the same underlying concept. We therefore also performed a principal component analysis with the maximum number of components restricted to one. Its solution is presented in the third column of table 4.8. The percentage of explained variance is now, of course, only 32.6% but all items (except EVGP7 again) have fairly high loadings on this factor.

Table 4.9. Principal component analysis of items on GP evaluation (n = 949): factor loadings

	TWO-COMPONENT SOLUTION (VARIMAX ROTATED)		ONE-COMPONENT SOLUTION
	FACTOR 1	FACTOR 2	FACTOR
EVGP 1	.736	-.053	.576
2	-.127	.717	-.580
3	-.343	.680	-.714
4	-.133	.721	-.587
5	.625	-.126	.544
6	-.168	.568	-.509
7	.200	.545	-.223
8	.623	.008	.452
9	-.280	.586	-.603
10	.665	-.268	.670
11	.599	-.338	.669
12	.659	-.115	.562
EIGENVALUE	3.91	1.62	3.91
% of VAR	32.6	13.5	32.6
CUMUL. %	32.6	46.1	32.6

In order to calculate the scores of the individuals on each of these factors, the factor coefficients presented in table 4.10 were used. By multiplying each item with its coefficient and adding up these products, factor scores for each individual are obtained for the 'inaccuracy perception (INACP2) and the accuracy perception (ACCURP2) from the two-component solution and for the inaccuracy perception (INACPT) based on the one-component solution. It is difficult to decide at this stage which of the two solutions or which factors should be retained for further analysis. We will therefore postpone the selection of the empirical counterpart for our theoretical concept until we have examined the relationships with the determining factors in the MIMIC model.

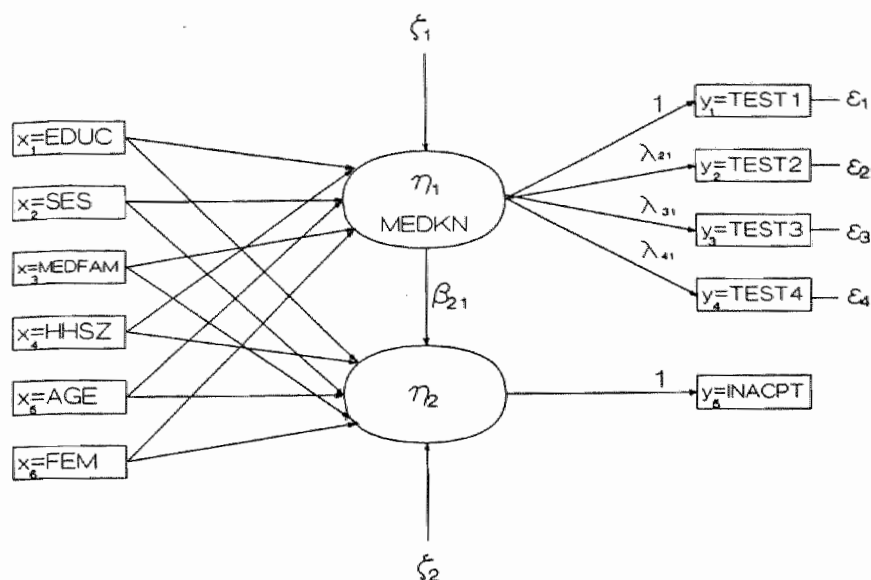
Table 4.10. PCA of GP evaluation items (n = 949):
factor coefficients

	TWO-COMPONENT SOLUTION		ONE-COMPONENT
	INACP2	ACCURP2	INACPT
EVGP 1	.301	.108	.147
2	.075	.300	-.148
3	-.023	.244	-.183
4	.073	.301	-.150
5	.241	.056	.139
6	.031	.226	-.130
7	.183	.281	-.057
8	.264	.115	.116
9	-.013	.213	-.154
10	.233	-.001	.172
11	.192	-.045	.172
12	.258	.067	.144

Because theory offers little indications as to what factors ought to determine this perception (in fact, θ_p is assumed exogenous in Chapter 3) we have used the same set of 'causes' as for the equation explaining medical knowledgeability. We also assume causality between knowledge and perception to run from the first to the second because knowledge will in general increase patients' ability to judge the accuracy of the information received. If indeed professional ignorance and uncertainty tend to be underestimated, as several authors have suggested (e.g. Pauly, 1978; Wennberg et al., 1982 or Breyer, 1982), then more information should lead, ceteris paribus, to a lower accuracy perception. The model specification used is given by the path-diagram in figure 4.3.

In table 4.11 the estimation results are presented of the structural equations for perceived accuracy for three MIMIC models, one for each principal component. The full model also includes the measurement

Figure 4.3. Path-diagram of MIMIC-information model



equations and structural equation for medical knowledge (η_1 = MEDKN) but these are not reported since the estimation results only differ sometimes in the third decimal from those presented in tables 4.6 (last column) and 4.7 (first column).

In fact, the reported equations are SIMIC models (ML regression equations) because 'perceived accuracy' is always measured by a Single Indicator and Multiple Causes. From the first column in table 4.11 we can see that the 'inaccuracy perception', as measured by all 12 items combined in one component INACPT, is positively affected by medical knowledge (η_1) and negatively by age. This implies that younger people and those with more medical knowledge have less confidence in the quality of GP services. All other background characteristics show insignificant effects and the proportion of explained variance is rather low (4.4%).

When the same model was estimated separately for each of the factor scores derived from the two-component-PCA, very different results were obtained. INACP2 was only positively related to MEDFAM at the 5% level of significance. This would imply that, controlling for other factors, only the individuals with medical family members had a significantly higher score for perceived inaccuracy of GP's. The explanatory power of this equation is very poor. On the other hand, a positive accuracy perception (ACCURP2) was found to be strongly positively determined by

Table 4.11. Estimation results of the MIMIC model in figure 4.2 for three definitions of perceived accuracy of GP

η_2	INACPT	INACP2	ACCURP2
$\eta_1 = \text{MEDKN}$.143 (2.2)	-.099 (1.5)	-.332 (5.1)
EDUC	-.012 (.24)	-.081 (1.7)	-.073 (1.6)
SES	.011 (.25)	-.004 (.09)	-.021 (.51)
MEDFAM	.049 (1.5)	.073 (2.2)	.007 (.23)
HHSZ	-.026 (.78)	.033 (.96)	.077 (2.4)
AGE	-.139 (3.8)	-.052 (1.4)	.153 (4.3)
FEM	-.046 (1.4)	.006 (.17)	.077 (2.5)
R^2	.044	.018	.144
$\chi^2(23)$	46.4	49.7	50.1

Note: The χ^2 test statistic refers to the full model and not only to the equation reported here.

age, female sex and household size and negatively by medical knowledge.

It has to be pointed out that the coefficients given in table 4.11 represent the **direct** effects of the background characteristics on the dependent variable. In order to find the **total** effects, the **indirect** effects (through η_1) have to be added. In most cases, the total effects are larger (in absolute value) than the direct effects which means that the indirect effect reinforces the direct effect.

All three fitted models are rejected by the χ^2 test with 23 degrees of freedom. We will select the INACPT variable as the measure of accuracy perception to be used in the full structural model because it has the lowest χ^2 test statistic (46.4) and because it summarizes all 12 EVGP items in one construct. In section 4.4 we will examine whether the predictions made in Chapter 3 regarding the effects of prior information Φ_i as measured by the medical knowledge latent variable, and of the perceived accuracy of GP information Φ_p as measured by the INACPT factor score, are compatible with our sample data.

4.4. A structural equation model of the demand for information and medical care

This section is divided into three parts. Before we present and discuss the estimation results, we will first describe the specification of the full model explaining the demand for care and information. The section is concluded with an investigation of the

robustness of the results with respect to some modifications of the basic model.

4.4.1. Specification of the demand equations

In the theoretical model outlined in Chapter 3, it was argued that especially a visit to a general practitioner at the patient's own initiative could be seen as largely consisting of a demand for advice. We will therefore concentrate in this empirical analysis on the explanation of the number of contacts with a GP initiated by the respondents during the observation period (GPCON) to test some of the predictions of the demand for information theory. As an alternative available to the patient we consider self-medication (SMED). For minor illnesses, consumers can use medicines without a prescription instead of - or before - going to see a doctor. As a measure of this sort of medical consumption, we will use the number of different types of self-medication taken during the nine weeks period.

In addition, we examine two types of prescribed medical consumption: the number of GP visits made on referral (GPREF) and the number of types of medicines used on prescription (FMED). These are clearly demands for medical care conditional upon the physician's advice which can be used to test for demand creation effects. We also hypothesized these demands to be interrelated in a hierarchical manner. Because self-medication is likely to be an option only prior to seeking professional advice, we included SMED in the GPCON and in the FMED equation in order to see whether it is used as a substitute for or complement of prescribed medication. The number of referred GP contacts is assumed to be dependent on the number of patient initiated contacts and the use of prescribed medicines is conditional upon both patient- and physician-initiated contacts.

The specification of the equations to explain these demands for care and information is guided by the variables included in table 3.1 of Chapter 3. The main explanatory variables of interest in this study are, of course, the variables measuring the two Bayesian components of consumer information Φ and θ_p . It is expected that both prior information and perceived inaccuracy will reduce the initial demand for information from general practitioners. Ceteris paribus, the opposite effects should occur for the demand for self-medication: more knowledgeable consumers and those with less confidence in general practitioner advice are expected to be more inclined to engage in 'self-doctoring'. If physicians are assumed to manipulate the accuracy of their advice in order to create additional demand for their services then again the possibilities for such manipulation should be more restricted for informed and/or sceptical consumers. We thus expect similarly negative effects of knowledge (as a proxy for Φ) and perceived inaccuracy $1/\theta_p$ (as a proxy for q) on the demand for repeat visits and prescribed medicines.

As we argued in Section 3.4.2, the initial health of the individual (H_0) is assumed to have a negative impact on b_0 , the consumer's prior estimate of the effectiveness of medical care. Although the comparative static effects of b_0 on the demand for information were shown to be non-linear and approaching zero at the extremes of b_0 , it seems reasonable to assume that it is positive in the range which is relevant for our sample of individuals. It is conceivable that the demand for information becomes negatively related to the expected marginal effectiveness of health care at very high levels of illness severity but such individuals are likely to be in hospital already and not participating in the Primary Health Care Survey. We can therefore expect a negative effect of H_0 on both the demand for information and the demand for medical care.

The question then arises how we should enter H_0 in the demand equations. In the section on health measurement we argued that it makes sense to distinguish between two components of an individual's general health status. In principle, one could expect that variations in both the permanent health state and the transitory flow of sick time may give rise to a demand for (information about) medical care. However, we will impose some additional restrictions on the direct effects of health on demand. It is plausible to assume that individuals will consult a doctor or take prescribed medicines at their own initiative only if they experience a deviation from their permanent health status, i.e. a transitory health loss. This means that permanent health influences this patient demand only indirectly through its impact on transitory health.

On the other hand, the use of prescribed medicines or GP services on referral may be a consequence of a (permanent) health problem that was present already before the beginning of this observation period. Permanent health therefore directly affects prescribed consumption of medical services. Transitory health can only indirectly exert an influence because the individual will have to consult a doctor³ first at least once. The hypothesized interrelationships between health and medical care utilization are summarized in figure 4.4. This recursive structure will be used in the basic model specification but in a later stage we will examine the validity of some of these model restrictions.

Because information and care are assumed to be normal goods, family income (FINC) should have a positive effect on their demand, ceteris paribus. However, as was pointed out before, due to the presence of extensive insurance coverage the impact of income level on medical consumption may be much weaker than for other goods and services. The same argument applies to the effect of money prices. We explained in section 3.4.2 that all sickfund beneficiaries paid the same fixed deductible amount for prescription drugs and general practitioner visits but there is one exempt category: the so-called WOPI-category (the widows, orphans, pensioners and invalids) paid a lower deductible

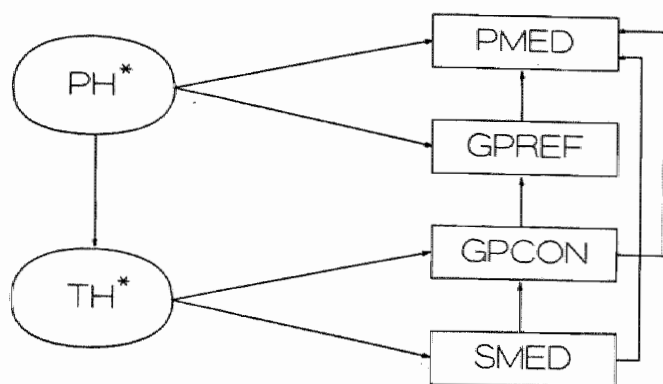


Figure 4.4. The relationships between health status and medical care utilization

for prescribed medicines and received full reimbursement of GP fees.⁴ It is well-known that these beneficiaries have a much higher medical consumption (see e.g. Dejardin and Praet, 1980). A dummy variable WOPI is included in the demand equations to test for possible differences in consumption between these two types of insurance beneficiaries, after correction for other differentials (health status, income, knowledge, etc.).

In order to adjust for differences in travel and time costs between different regions, we have added one or two indicators of health care supply availability: the pharmacy (PHARD) and general practitioner density (GPDENSI) in the municipality of residence of the respondent. In the equation for GP contacts on referral, the coefficient of this variable may also represent to some extent an effect of supplier-induced demand (Pauly, 1978). In regions with high provider densities average fee-for-service income is lower unless the suppliers are able to induce additional demand (e.g. repeat visits) for their services.

Finally, four background characteristics (AGE, FEM, EDUC and MEDFAM) have been entered into the demand equations to control for a possible direct influence apart from their indirect effects through health and information. The MIMIC measurement models for health and information are almost identical to the ones estimated and presented in sections 4.2 and 4.3. The only difference is that we have here assumed that knowledgeability may exert a direct positive influence on health status, irrespective of consumption behaviour (e.g. through healthier habits). On the other hand, individuals with a lower permanent health status, and therefore probably higher levels of health care utilization in the past, may have acquired more knowledge about health and medical care. In the structural model estimation, we have tested for this simultaneity by not restricting the coefficients of MEDKN in the

PH and TH equations, and that of PH in the MEDKN equation to be zero.

4.4.2. Estimation results

The total linear structural equation model comprises eight structural equations and fifteen measurement equations. The LISREL equation (4.2) now becomes

$$y = \Lambda \eta + \epsilon \quad (4.2)$$

where $y' = (\text{SMED, GPCON, GPREF, FMED, CCOND, PGPC, COMPL, LRAD, LNAD, LDISAB, TEST1, TEST2, TEST3, TEST4, INACPT})$

$$\eta' = (\text{SMED, GPCON, GPREF, FMED, PHEALTH*}, \text{THEALTH*}, \text{MEDKN*}, \text{INACPT})$$

Note that the first four and the last η -variable are by definition identical to the directly observable y -variables. The corresponding measurement equations reduce to identities with λ equal to one and the e -variance equal to zero: $\eta_1 = y_1$, $\eta_2 = y_2$, $\eta_3 = y_3$, $\eta_4 = y_4$, and $\eta_8 = y_{15}$.

The structural equation part (4.7) now equals

$$\eta = B\eta + \Gamma x + \zeta \quad (4.7)$$

with η = as above

$$x' = (\text{EDUC, SES, MEDFAM, HHSZ, AGE, FEM, PSPP, LIHIST, FINC, WOPI, PHARD, GPDENS})$$

This model is identified because it can be split into three parts which are identified each (see e.g. Wiley, 1973). A formal proof of the identification of this model is provided in appendix 4D.

As input for the maximum likelihood estimation procedure the correlation matrix was used. It is worth noting from the correlation matrix given in appendix 4B that the simple correlations between the knowledge test scores and the medical consumption variables of interest are fairly low but positive for SMED, zero for GPCON and negative for GPREF and FMED. The correlations between INACPT and the demand variables are similarly low but all negative. We will see what happens to these associations when other influences are controlled for in the causal model structure.

Table 4.12 presents the estimation results (λ -coefficients, t -values and reliability coefficients) for the measurement equations of the three latent variables. Comparison with table 4.4 (model 1D) and table 4.6 (MIMIC model) learns that integration of these measurement equations in a larger structure leaves the estimates virtually unaffected. If any change at all can be detected it is a slight improvement of the t -values and/or the ρ -coefficients. This stability result is reassuring in view of the efforts that were made to obtain adequate measurements of the theoretical constructs.

Table 4.12. Maximum likelihood estimation results of the LISREL measurement model for the three unobservables $y = \lambda\eta + \epsilon$ ($n = 949$)

	CCOND (η_5)	PGPC (η_6)	COMPL (η_7)	$\hat{\theta}_{665}$
FHEALTH (η_5)	-1.0 (-) .247	- 1.094 (12.6) .296	-1.412 (13.4) .492	.088 (3.0)
	LRAD (η_8)	LRAD (η_9)	LDISAB (η_{10})	$\hat{\theta}_{698}$
THEALTH (η_6)	- .776 (18.3) .318	- .776 (18.3) .318	-1.0 (-) .527	- .09 (3.5)
	TEST1 (η_{11})	TEST2 (η_{12})	TEST3 (η_{13})	TEST4 (η_{14})
MEDKN (η_7)	1.0 (-) .565	.902 (17.3) .460	.894 (17.2) .452	.419 (8.5) .099

Note: λ -coefficient, absolute t-value in parentheses, and reliability coefficient (ρ)

In table 4.13, the LISREL estimates of the structural part of the model are presented. The first four columns contain the unstandardized β - and γ -coefficients (and their t-values) for the health care demand equations. If we go through the results from the top to the bottom of the table, we see first of all that these different demands are strongly interrelated. Self-medication (SMED) appears to have a negative effect on both the frequency of seeing a GP and using prescribed medicines. This probably reflects the fact that these over-the-counter drugs are often used for less severe health problems.⁵ Perhaps surprisingly, the number of GP contacts at the initiative of the patient has a negative influence on the number of GP contacts on referral. This suggests that most of the referred contacts result from earlier contacts **before** the period of observation rather than from patient-initiated contacts during this period. It may also imply that these referred contacts occur mainly for more chronic patients under longer term treatment. Finally, both patient- and doctor-initiated contacts increase the use of prescribed drugs considerably. This is not surprising since pharmacotherapy is the far most common type of intervention used in Belgian general practice.⁶

As in most studies of health care utilization, the impact of the health variables dominates all other influences. Transitory health in the patient-initiated and permanent health in the doctor-initiated utilization equations show large and strongly significant negative

Table 4.13. Maximum likelihood estimation results of the LISREL model: $\eta = B\eta + \Gamma\zeta + \xi$ ($n = 949$)

	η_1 SPED	η_2 GFOON	η_3 GPREF	η_4 FMED	η_5 FHEALTH	η_6 THEALTH	η_7 METKN	η_8 INACT
η_1 SPED								
η_2 GFOON	-.095 (2.9)							
η_3 GPREF		-.127 (3.3)						
η_4 FMED								
η_5 FHEALTH			-.746 (7.0)	-1.409 (11.6)		1.271 (11.8)	-.155 (1.1)	
η_6 THEALTH	-.487 (9.4)	-.798 (13.7)						
η_7 METKN	.127 (2.1)	.121 (2.0)	-.136 (2.1)	-.011 (.23)	.017 (.21)	.034 (.65)		.138 (2.1)
η_8 INACT	-.022 (.73)	-.011 (.37)	.023 (.73)	-.012 (.56)				
x_1 EDUC	.100 (2.4)	-.032 (.80)	-.027 (.62)	-.009 (.27)	-.029 (.76)	.070 (2.0)	.243 (7.0)	-.011 (.23)
x_2 SES							.206 (6.5)	.014 (.31)
x_3 MEDFAM	-.016 (.50)	-.042 (1.4)	.021 (.64)	-.006 (.23)	-.004 (.18)	-.029 (1.1)	.075 (3.1)	.050 (1.5)
x_4 HSHZ					.068 (3.7)	-.028 (1.2)	.023 (.83)	.026 (.76)
x_5 AGE	-.084 (2.2)	.002 (.06)	-.015 (.36)	.021 (.64)	-.118 (5.1)	.101 (3.2)	-.083 (2.5)	-.138 (3.7)
x_6 FEM	.144 (4.6)	.023 (.76)	-.019 (.56)	-.036 (1.3)	-.127 (5.9)	.094 (3.4)	.042 (1.4)	-.048 (1.5)
x_7 PSFR					-.068 (3.9)	-.086 (3.8)		
x_8 LIHST					-.079 (4.4)	.046 (2.0)		
x_9 FINC	.007 (.18)	-.037 (1.1)	-.017 (.46)	.015 (.60)				
x_{10} WOPI	-.006 (.17)	.070 (2.0)	.011 (.29)	.029 (1.1)				
x_{11} HARD	.008 (.25)			-.028 (1.3)				
x_{12} GPENS		.000 (.002)	.040 (1.3)	-.011 (.52)			-.021 (.88)	-.034 (1.1)
R^2	.185	.336	.122	.705	.201	.703	.384	.044

$\chi^2_{(196)} = 933$
 $AGFI = .881$

direct influences on demand.

The effects of the variables in which we are primarily interested in this study, MEDKN and INACPT, are much smaller. Inaccuracy perception of general practitioners, as it is measured by the factor score INACPT based on 12 items, shows no significant effect on utilization at all which makes it not worthwhile to look at the signs of the coefficients. Medical knowledgeability (MEDKN), on the other hand, has a positive coefficient for SMED and GPCON (both significant at the 5% level), a negative coefficient for GPREF (significant at the 5% level) and a non-significant coefficient for FMED. Two of these effects, namely on SMED and GPREF, are consistent with our theoretical predictions but the major prediction of this theory, i.e. that more knowledgeable consumers should, *ceteris paribus*, demand less frequently information from their GP is clearly rejected. It is even so that the opposite result seems to hold. A more elaborate discussion and interpretation of these effects is deferred to the conclusion of this chapter which also incorporates the results of a robustness analysis of some model modifications.

The effects of the exogenous x-variables on utilization are in general small and, with some notable exceptions, non-significant. Family income (FINC) is never significant and the insurance dummy (WOPI) indicates that this category of insurance beneficiaries only have a significantly higher number of patient-initiated contacts with the GP. If health status, income, age and other differences are adequately controlled for in this analysis, then this finding could be interpreted as a pure price effect. It may, however, not only be a money price effect (GP services were free for these people) but may also reflect a time price effect: pensioned and disabled people can be expected also to face lower time costs when visiting a GP.

Neither the number of pharmacies (PHARD), nor the number of general practitioners (GPDENS) per capita show a statistically significant influence on the frequency with which their services are being used by the sample population. However, it deserves mentioning that the only consumption category for which one could expect supplier-induced demand (over and above the time price effects of shorter distances and waiting times) because it is care prescribed and delivered by the same provider, i.e. physician-initiated referral contacts, also shows the largest availability effect. The implied elasticity at the means of GPREF and GPDENS equals .46.⁷ Although it is statistically not significantly different from zero it deserves some closer investigation in future work. The non-significance may simply be a result of limited power of the test due to our limited sample size in which only 10% of the individuals report any referred contacts. Rossiter and Wilenski (1981) for example, who analyzed the 1977 VS National Medical Care Expenditure Survey ($n = \pm 40,000$), also found that physician density had a small but significant positive effect on physician-initiated but not on patient-initiated visits. Evaluated at

the mean, they estimated on elasticity of .06.

For the remaining exogenous (taste) variables in the demand equations (EDUC, MEDFAM, AGE, FEM) only some direct influence on SMED was found: women and higher educated people use more and older people fewer non-prescribed medicines. The direct effects of these variables on the other types of demand were statistically insignificant and therefore negligible compared to their indirect effects through health and knowledge.

When we turn to the structural equations presented in the last four columns of table 4.13, we see again that the integration of the sub-models into the larger model, estimated in the measurement sections 4.2 and 4.3 has had little effect on the magnitude and sign of the estimated coefficients. We will therefore focus our attention on the added simultaneous relationship between health and medical knowledge. The direct effects of MEDKN on permanent and transitory health are positive but not statistically significant. This is remarkable because education does have a significant positive influence on transitory health which implies that, *ceteris paribus*, the higher educated are less likely to report sick time. In contradiction to what one would expect, general knowledge capital seems more beneficial to one's health than specific medical knowledge capital. This finding cannot be due to the fact that people with more extensive illness experience also acquire more knowledge about health and medical care because this reverse causality has been introduced in the MEDKN equation: indeed a negative coefficient is estimated for PH but it is not statistically significant. We must conclude that the cross-sectional association between knowledge and health is weak when other influences (e.g. education) were controlled for.

The squared multiple correlations for the structural equations are calculated as $R^2_i = 1 - \text{var}(\xi_i)/\text{var}(\eta_i)$ (Jöreskog and Sörbom, 1981, p. I.37). They indicate a reasonably good fit for most of the relations except the equation for INACPT. Explained variance is highest (about 70%) for prescribed medicines and transitory health and lowest for self-medication and referred contacts with the GP (less than 20%). The χ^2 test statistic value for the total model (measurement and structural part together) indicates rejection of the null hypothesis that the correlations between the observed variables have been generated by this model. We have emphasized before that the sample size and the non-normal distribution of several of the dependent variables tend to increase the value of the test statistic. According to some authors, our fit is not such a bad result at all with a sample size of about 1000: "We judge a ratio (test statistic divided by the number of degrees of freedom) of around 5 as beginning to be reasonable, based on our experience in inspecting the sizes of the residuals which accompany varying χ^2 values" (Wheaton et al., 1977, p. 99). Since this ratio equals 4.76 (= 933/196) for our model, we can claim some reasonability as to the goodness of the fit despite formal rejection by the

χ^2 test. The goodness-of-fit index adjusted for degrees of freedom (AGFI) equals .881 and the total coefficient of determination for the structural model (R^2_{η}) is .618. Both measures are difficult to judge at face value but will be used for comparison with alternative model specifications in section 4.4.3.

Finally, it is worthwhile to examine the total effects of each variable on the endogenous variables in the model. Table 4.13 only reports the estimated direct effects but a lot of variables also exert indirect influences which have to be added in order to obtain the total effects. Furthermore, a meaningful comparison can only be made for the standardized estimates. These are given in table 4.14 for a selected number of η - and x-variables.

Some interesting findings emerge from this exercise. First of all, the indirect effects of FHEALTH on SMED and GPCON are much more important than the indirect effects of THEALTH on GPREF and FMED. The latter is even positive because THEALTH reduces GPCON and thereby increases GPREF. Note also, that, as a consequence of the simultaneous determination of FHEALTH and MEDKN, total effects of these variables on themselves are produced! Such "cycling effects" are calculated as the sum of an infinite geometric series (see Jöreskog and Sörkom, 1981, p. III.92) which is equal to $\beta_1\beta_2/(1 - \beta_1\beta_2)$, where β_1 and β_2 represent the mutual total effects of the two η 's (FHEALTH and MEDKN). It can be seen that the series converges, at the value of -.003, indicating stability of the system.

Furthermore, the positive direct effects of MEDKN on SMED and GPCON are larger than the total effects. The indirect effects of MEDKN via SMED, FHEALTH and INACPT reduce the total effect to about half the direct effect. This means that more self-medication, slightly better health and a less favourable perception of general practitioner services reduce the partial positive impact of medical knowledge on patient-initiated GP contacts considerably. The remaining effect is still positive. Thus we reach the same unexpected result as Hay and Leahy (1982) but nevertheless we can reject their interpretation of it. They argue: "It is certainly plausible to interpret these results as consistent with the view that medically informed people have more faith in medicine, and utilize services with little true marginal effect, but this begs the question of who is really medically informed." (p. 242) This explanation is in contradiction with our finding of a negative indirect effect of MEDKN on GPCON through INACPT, which is, however, much smaller than the direct positive effect: they seem to consult the GP more often despite the fact that they show less confidence in GP's in general.

We can observe once more that the health gains of both MEDKN and EDUC are small and that the signs of their effects on the demand for medical care are the same except for GPCON. AGE and FEM have substantial total effects on consumption due to their significant

Table 4.14. Standardized total and direct effects of some η and x variables on the η -variables

A. TOTAL EFFECTS		η_5 PHEALTH	η_6 THEALTH	η_7 MEDKN	x_1 EDUC	x_5 AGE	x_6 FEM
η_1	SMED	-.635	-.487	.097	.108	-.063	.184
η_2	GPCON	-.966	-.752	.066	-.051	.043	.065
η_3	GPREF	-.600	.096	-.153	-.035	.073	.057
η_4	FMED	-1.585	-.087	-.052	-.009	.218	.134
η_5	PHEALTH	-.003	0	.017	-.025	-.119	-.126
η_6	THEALTH	1.263	0	.055	.047	-.052	-.065
η_7	MEDKN	-.155	0	-.003	.247	-.064	.062
η_8	INACPT	-.021	0	.137	.023	-.147	-.039
B. DIRECT EFFECTS		η_5 PHEALTH	η_6 THEALTH	η_7 MEDKN	x_1 EDUC	x_5 AGE	x_6 FEM
η_1	SMED	0	-.354	.096	.100	-.084	.144
η_2	GPCON	0	-.581	.091	-.033	.002	.023
η_3	GPREF	-.370	0	-.102	-.027	-.015	.019
η_4	FMED	-.698	0	-.009	-.009	.021	.036
η_5	PHEALTH	0	0	.026	-.058	-.237	.256
η_6	THEALTH	.870	0	.035	.097	.139	.129
η_7	MEDKN	-.103	0	0	.323	-.110	.056
η_8	INACPT	0	0	.103	-.011	-.138	.048

indirect effects via permanent and transitory health. The elderly seem to be particularly intensive users of prescribed medicines and younger people take more non-prescription drugs whereas women use more of both types.

Before drawing up some general conclusions with regard to our findings, we will first investigate the robustness of our model estimates with respect to some model modifications in the next section.

4.4.3. Robustness analysis

The estimation results presented in the previous section brought out that some of the major predictions of our theoretical model were not confirmed or even had to be rejected. Because the estimated coefficients on which these tests were based often showed substantial standard errors and the t -values were therefore sometimes close to the critical levels of statistical significance, it is worthwhile to examine how robust these results are, i.e. how sensitive they are to some changes in (a) the sample choice, and (b) the model structure.

4.4.3.1. Sample selection

First, we will consider the question whether the sample choice is adequate. Until now we have analyzed the entire sample of adults whereas in fact, strictly speaking, the theory of the demand for diagnostic and therapeutic information only applies to individuals who experience their initial health status H_0 as suboptimal and therefore consider whether or not to undertake action or seek advice on what to do. About one quarter of the sample did not report any health complaint during the whole observation period and consequently had little reason to contemplate a visit to their GP. Only about 3% of the non-complainers reported a GP contact. When this upper 25% of the health distribution is deleted from the sample, the remaining 720 individuals, who have reported at least one health complaint, represent a somewhat less healthy truncated subsample.⁸ The majority of 'complainers' are still non-users of medical care. By excluding the individuals without health complaints, we avoid the problem that patterns of utilization for preventive reasons may differ from those for curative purposes. For instance, it is not impossible that medically informed people consult a doctor more for preventive and less for curative reasons (see e.g. Hay and Leahy, 1982).

When the basic model was reestimated on the basis of the correlation matrix for this sub-sample, the results presented in table 4.15A and B were obtained. In general, the deletion of a large part of the healthier individuals have weakened a lot of the observed correlations. Most of the estimated coefficients have the same sign as in the basic model but both the magnitude and statistical significance (as indicated by the t-values) is often reduced. For example, the effects of medical knowledge (MEDKN) on SMED, GPCON and GPREF are virtually the same as before but are now only significant at the 10% level. Although the coefficients of determination of the structural equations on average do not improve, the χ^2 test statistic indicates a substantially superior fit of the model to these data. Clearly, this is mainly a consequence of the reduction of the sample size but even the AGFI, which is independent of the sample size, indicates a slightly better fit. Some of the improvement in the goodness of fit measures may be due to the increased reliability coefficients of all health indicators, except of course COMPL, in the measurement part of the model.

Of primary importance for our purposes, however, is the stability of the major results: the effects of consumer information on the demand for care and information are the same for the sub-sample of individuals with some perceived morbidity as for the sample as a whole.

	η_1 SMED	η_2 GPOON	η_3 GPREF	η_4 PMED	η_5 PHEALTH	η_6 THEALTH	η_7 MEDKN	η_8 INACT
η_1 SMED	-0.44 (1.2)							
η_2 GPOON		- .085 (3.1)						
η_3 GPREF			- .128 (3.1)					
η_4 PMED				.092 (2.8)				
η_5 PHEALTH					- .712 (5.9)			
η_6 THEALTH						1.053 (9.0)		
η_7 MEDKN							- .128 (.73)	
η_8 INACT								.250 (3.2)
x_1 EDUC	.079 (1.6)	-.049 (1.0)	-.016 (.31)	-.011 (.25)	-.007 (.14)	.073 (1.6)	.240 (6.1)	-.078 (1.4)
x_2 SES							.225 (6.1)	.002 (.04)
x_3 MEDFAM	-.011 (.30)	-.034 (.96)	.026 (.65)	-.004 (.13)	-.007 (.26)	-.021 (.67)	.074 (2.6)	.049 (1.3)
x_4 HSEZ					.086 (3.5)	-.031 (.99)	.039 (1.2)	-.042 (1.1)
x_5 AGE	-.105 (2.3)	.011 (.25)	-.079 (1.4)	-.028 (.60)	-.213 (6.7)	.197 (4.5)	-.089 (1.8)	-.147 (3.4)
x_6 FEM	.143 (3.4)	.025 (.72)	-.012 (.30)	-.018 (.55)	-.103 (3.7)	.095 (2.7)	.045 (1.4)	-.045 (1.2)
x_7 PSFR					-.065 (2.9)	-.096 (3.3)		
x_8 LIHIST					-.057 (2.5)	.050 (1.7)		
x_9 FINE	.016 (.38)	-.042 (1.1)	-.011 (.25)	.031 (1.0)				
x_{10} WOPI	-.009 (.20)	.074 (1.8)	.001 (.02)	.017 (.54)				
x_{11} PHARD	.013 (.36)			-.029 (1.1)				
x_{12} GPENS		-.003 (.09)	.053 (1.5)	-.013 (.48)			-.017 (.61)	-.015 (.42)
R^2	.083	.262	.136	.649	.303	.527	.383	.057

$\chi^2_{(196)} = 678$
AGFI = .896

Table 4.15B. Maximum likelihood estimation results of the LISREL measurement equations $y = \lambda\eta + \epsilon$ for subsample (n = 720)

	Y5 COOND	Y6 PGPC	Y7 COMPL	cov(ϵ_5, ϵ_6) $\theta_{\epsilon 65}$
η_5 RHEALTH	-1.0 (-) .312	-1.001 (11.0) .312	-.845 (9.7) .223	.002 (.07)
	Y8 LRAD	Y9 LRAD	Y10 LDISAB	cov(ϵ_8, ϵ_9) $\theta_{\epsilon 98}$
η_6 THEALTH	-.791 (14.8) .343	-.791 (14.8) .343	-1.0 (-) .547	-.150 (4.7)
	Y11 TEST1	Y12 TEST2	Y13 TEST3	Y14 TEST4
η_7 MEDKN	1.0 (-) .574	.872 (14.8) .436	.885 (15.0) .449	.411 (7.4) .097

Notes: see table 4.1

4.4.3.2. Model structure

Secondly, we considered the possibility that the results of our model might be very sensitive to some specification changes. In going from the theoretical model to this particular empirical specification tailored to the dataset at our disposition, some ad hoc assumptions had to be made with regard to which variables ought to be included in which equations. Some of the arbitrary choices that were made in this process, often on the basis of intuitive plausibility, may have had a substantial impact on the outcomes. Inevitably, in a structural equation model as large as the one presented here quite a number of restrictions on the structure have to be imposed, not all of which can be justified from the theoretical basis. We have investigated two model modifications which may be expected to potentially affect the coefficients of the MEDKN and INACPT coefficients.

First, we investigated the effects of deleting the education variable from all demand and health equations. In the demand equations, it has no other justification than as a control variable for 'taste' differences among people with different levels of education. But because general knowledge and specific medical knowledge are clearly strongly related, collinearity with MEDKN may affect the standard errors of the coefficients.

By restricting the coefficients of EDUC to zero in the first six structural equations (γ_{11} to γ_{61} in table 4.13) we assume that education only influences medical care utilization and health indirectly through MEDKN and INACPT. Some of the results of this more restricted model estimation are presented in table 4.16. Only the re-estimated coefficients β_{17} to β_{67} for MEDKN and β_{18} to β_{68} for INACPT are reported because all others were only marginally different from table 4.13 and therefore did not need to be replicated here.

Table 4.16. Selected results for model reestimation with six education coefficients restricted to zero

	MEDKN η_7	INACPT η_8	R^2
η_1	.216 (4.1)	-.025 (.81)	.183
η_2	.104 (2.1)	-.010 (.34)	.335
η_3	-.164 (3.0)	.024 (.75)	.123
η_4	-.021 (.48)	-.013 (.59)	.705
η_5	-.022 (.49)	0	.209
η_6	.100 (2.3)	0	.699

$$\chi^2_{(202)} = 944$$

$$AGFI = .882$$

From comparison with table 4.13 we notice that the effects are in line with what one could expect. In general, the INACPT coefficients remain insignificant, but the statistical significance of the MEDKN coefficients improves as a result of the reduced collinearity. The effects of MEDKN on SMED and THEALTH have increased substantially because they now also capture the (significant) indirect influence of education. However, none of the previous conclusions regarding the impact of prior knowledge on demand is changed by this model modification. The increase of the χ^2 test statistic equals $944 - 933 = 11$ and is associated with a gain of six degrees of freedom. It exceeds the 10% critical level (10.6) but not the 5% critical level (12.6). Apparently, the fit improvement (due to deletion of four insignificant coefficients) by and large compensates the fit deterioration (due to the zero restrictions on two significant coefficients). On the whole, there seems to be no reason to prefer this model specification to the basic one because the main effects of interest remain unaltered and the original specification allows to separate out the general schooling and specific medical knowledge effects.

The second model modification which we considered produced more

dramatic effects. When we wanted to test the validity of our assumptions that the direct effects of PHEALTH on patient-initiated demand (SMED and GPCON) and those of THEALTH on physician-initiated demand are indeed zero, we ran into problems of convergence of the LISREL iteration procedure. The LISREL program could not reach an optimal solution within the normal time limit. This non-convergence problem is most likely to be due to the presence of a high degree of multi-collinearity between the two latent health variables.⁹ The correlation between PHEALTH* and THEALTH*, as specified in the basic model, was .807. Although we showed in section 4.2.2 that this is still significantly different from 1, it is relatively high for a correlation coefficient based on a sample of some thousand individuals. To compare, the highest correlation among the observed variables is .633 between education (EDUC) and socio-economic status measured by some ordering of occupations (SES). Consequently, it becomes extremely difficult to disentangle the separate effects of both health variables in the same equation. Because we were forced to use no more than one health component variable in the demand equations, we chose to stick to our original specification.

4.5. Conclusions and discussion

In this chapter an attempt was made to test empirically some of the predictions of the theoretical model presented in the previous chapter. For that purpose, a structural equation model was developed to explain two patient-initiated demands for medical care (self-medication and GP contacts) and two physician-initiated demands (referred GP contacts and prescribed drug use) of a sample of nearly a thousand adults drawn from the general Flemish population. Individual health status was measured as a latent vector consisting of two components: a permanent stock of health and a transitory flow of healthy time. In a similar way, two Bayesian concepts of consumer information about medical care in general, and about primary care provided by general practitioners in particular, were distinguished: a MIMIC index for medical knowledgeability was used to measure prior information and a principal component was derived to measure the respondents' perception of the accuracy of GP advice. The reliability and validity of each of these constructs was examined by means of the LISREL model. The advantage of this MIMIC approach was shown to be that there is no longer a need to use determinants ('causes') or observable consequences ('indicators') as proxies for the relevant 'true' concepts in the analysis. Instead, several indicators are combined into one latent factor in order to increase the reliability by correcting for measurement errors. The validity of these indices was studied by relating them to a number of explanatory variables. The results of these measurement submodels proved to be very robust with respect to integration into the larger structural model. This provides an important justification for the stepwise strategy which we

adopted, of sorting out the measurement problems before performing any hypothesis testing in the total structural model. Especially for large and complex models this procedure seems fruitful and computertime-saving.

With respect to the tested predictions of the Bayesian model of demand for information when product quality is uncertain, the following conclusions can be drawn. First of all, better medically informed individuals do not initiate fewer visits to their GP. On the contrary, *ceteris paribus*, they report significantly more of such contacts, although the effect is small. The total effect of medical knowledge-ability is smaller than the direct effect, mainly because it has also a positive influence on the use of self-medication, which acts as a substitute for GP contacts.

The opposite effect was found for physician-initiated GP contacts: people with more medical knowledge had significantly less GP contacts on referral and the direct effect is reinforced by some indirect effects. No influence could be detected of prior information on the use of prescription drugs. Moreover, none of these demands was significantly affected by our measure of the (in)accuracy perception of GP diagnostic and therapeutic information. Estimation of the basic model on a subsample from which the 'healthy' individuals were deleted and some re-estimations with slightly modified specifications indicated that these results were fairly robust.

On the basis of these results, we have to conclude that neither for θ_p nor for ϕ the analysis confirmed the predicted effects on the demand for information as measured by the number of GP contacts at the patient's initiative. The inconsistency of these results with some of the major theoretical predictions stands in sharp contrast to their consistency with earlier findings in the literature on the demand for health care. Investigators who did include proxies for medical knowledge in their empirical demand functions either found no significant effect (e.g. Manning et al., 1982; Wagstaff, 1980; Hay and Leahy, 1982 (Tobit function coefficient) or a positive coefficient (e.g. Hay and Leahy, 1982, OLS-coefficient). In all cases, if any effect was found, it was positive but small. Only one study included a variable which is similar to our INACPT variable: Manning et al. (1982) used a four-item scale to measure a 'favorable attitude toward the efficacy of doctors and medical services'. It had a non-significant negative effect in a reduced form equation explaining (the logarithm of) outpatient expenditures for medical care of the adults which participated in three sites of the Rand Health Insurance Study.

Since all these results were based on larger sample sizes than ours we can have some confidence that our finding of a number of insignificant effects is not simply due to a lack of test power. Moreover, our structural modelling approach yields more insight into the causal processes that may have generated these results than the single equation reduced form approaches. For example, Hay and Leahy (1982)

are not sure whether their finding of a positive effect of the MEDFAM (family member in medical or health service occupation) dummy variable is not due to better access ('professional courtesy office visits') or 'more faith in medicine' of these people. Our approach enables us to rule out these possibilities of spurious effects because we could include both medical knowledge and its hypothesized causes and consequences in the GPCON equations. We found that, when controlling for the positive effect of consumer prior information, MEDFAM and INACFT had (non-significant) negative effects on the demand for contacts.

A second advantage of our analysis is that we could distinguish between patient- and physician-initiated consumption and this distinction turned out to be crucial. For physician-initiated contacts, indeed the effect of medical knowledgeability is significantly negative and even larger than the positive effect on patient-initiated demand. Although this finding is consistent with Pauly's (1980) hypothesis that prior information effectively limits the provider's ability to induce demand, it does not necessarily rule out other possible explanations. For instance, more informed GP clients may show better compliance with prescribed therapies and therefore need less repeat visits. However, in combination with the consistently positive, though non-significant, effect of general practitioner density in the place of residence, the existence of physician-induced demand for primary care through the generation of repeat visits, remains an interesting topic for further research. Only the analysis of larger datasets, which include more observations on this patient-initiated consumption, and preferably complemented with similar information on the use of (in Belgium directly accessible) medical specialists, can shed more light on this issue. But even when good data are available, the identification of supplier-induced demand by means of econometric analysis will remain a subject of controversy and debate among health economists.¹⁰

Does all this mean that the theory of consumers (patients) as Bayesian decision-makers, who demand information about a product with uncertain quality, can be rejected as inappropriate for the demand for medical care? Our results indeed suggest that prior information (and thus uncertainty) has only weak linear effects on demand behaviour. Health status dominates virtually all other influences, including the consumer's perception of the quality of GP information. Our conclusion is therefore that the demand for information theory, despite its theoretical appeal, contributes little to a better empirical explanation of the quantity of health care demand. Qualitative effects could not be tested adequately with the available dataset. As was pointed out already in Section 4.3.1, medical knowledgeability may manifest itself by more appropriate demand behaviour rather than in more or less overall utilization. If the level of consumption of the more knowledgeable consumers can indeed be considered as more appropriate, then our results suggest that the poorly-informed under-

utilize self-medication and general practitioner services. It should also not be forgotten that the effects of information may be biased by the institutional obligation to seek professional advice in order to obtain a drug prescription, or a sickness absence certificate. As a result, some proportion of the GP contacts may be "unnecessary" in the sense that patients knew in advance how to interpret and act upon certain symptoms but were nevertheless legally required to contact their GP. Since this proportion is likely to be higher for more informed patients, any negative effects of prior information on utilization of GP services tend to be reduced by this requirement.

As a by-product of the structural model analysis, we also tested whether medical knowledgeability had any direct effects on individual health status. When educational level was controlled for, no significant influence of MEDKN could be estimated. Education only had a small positive (linear) effect on transitory health. This evidence contrasts sharply with the strong partial schooling-health correlations which are often found in individual cross-section studies (see Chapter 2). Moreover, it indicates that if schooling has a positive influence on health, it is not primarily because the better-educated are more knowledgeable about health and medical care matters, but rather because they perhaps enjoy more favourable environmental circumstances (occupation, hygiene, housing, etc.). However, this conclusion is very tentative given the data limitation and the rather ad hoc specification of the health equations in this model. In the second part of this thesis, a more detailed analysis of the education-health relationship is envisaged by making use of the advantages offered by longitudinal data.

NOTES

1. In the meantime, cost-sharing has been substantially extended and modified with proportional coinsurance rates (see e.g. Van Doorslaer, 1984).
2. Another possibility is to adjust the power of the test, i.e. the probability to reject an incorrect model (see e.g. Saris et al. (1985), or Saris et al. (1986)).
3. General practitioner is too narrow here because also the medical specialists are directly accessible without a referral in Belgium.
4. In 1976, the deductible amounts were 60 Bfrs for prescribed drugs and 40 Bfrs for a GP visit, whereas the WOPI's only had to pay 35 Bfrs for prescription drugs.
5. Prescribed and non-prescribed use of these medicines according to the different drug categories is described extensively in Foets et al. (1985).
6. Nuyens (1980) reports that 78% of all GP contacts result in pharmacotherapy (p. 95, table 8).
7. The elasticity is calculated as $.457 = .013 \times \frac{5.7}{.162} = \gamma \times \frac{\text{mean GPDENS}}{\text{mean GPREF}}$, where γ is the coefficient of GPDENS in the GPREF equation when the covariance matrix of the observed variables was used as the input matrix. The γ 's can then be interpreted as normal regression coefficients. This is not so for the coefficients obtained with the correlation matrix as input.
8. A side effect of this truncation is that the skewness of the distribution of the dependent variables in our model (see appendix A) is somewhat reduced and therefore closer to the normality ideal.
9. According to Wagstaff (1986), who encountered similar problems, the degree of arithmetic precision required to ensure convergence becomes extremely high when there is a high degree of multicollinearity among the explanatory variables.
10. Cfr. the recent editorials in the Journal of Health Economics by Reinhardt (1985), Phelps (1986) and Fuchs (1986).

Part II. EDUCATION AND HEALTH

Chapter 5. Economic analysis of the schooling-health relationship

"The concept of human capital ... is the idea that people spend on themselves in diverse ways, not only for the sake of present enjoyments but also for the sake of future pecuniary and non-pecuniary benefits" (Blaug, 1980, p. 225)

5.1. Introduction

The first part of this thesis is concerned with an exploration of the direct linkage between consumer information and the demand for medical care. It provides an examination of one specific mechanism, i.e. Bayesian decision making, through which human capital, in the form of prior knowledge, might affect medical consumption and thereby health status. In this second part we take a broader view to investigate the overall effect of one type of human capital, i.e. education, on the other type of human capital, i.e. health. This chapter deals with the theoretical and empirical problems which are associated with an econometric analysis of the schooling-health relationship. But first, we will give an indication of the importance and relevance of the issue, both from a theoretical and a health policy point of view.

Probably the most comprehensive economic analysis of this relationship has been performed, again, by Grossman (1975). He argued: "Several recent studies in the United States indicate that among socio-economic variables, years of formal schooling completed is probably the most important correlate of good health ... This finding emerges whether health levels are measured by mortality rates, morbidity rates, or self-evaluation of health status, and whether the units of observation are individuals or groups." (Grossman, 1975, p. 147). Numerous European studies suggest that a similarly strong correlation also holds outside the U.S.A. (for a recent survey of socio-economic inequalities in health, see Mackenbach and Van der Maas, 1987). Furthermore, the statistical relationship usually remains strongly significant when other covariates, such as income, are controlled for.

These findings have led some authors (e.g. Auster, Leveson and Sarachek, 1969) to suggest that the rate of return on increases in health via higher expenditures on education may far exceed the rate of return on increases in expenditures for medical care. Others, like Haveman and Wolfe (1984), have argued that standard estimates of earnings differentials due to incremental schooling have seriously underestimated the total benefits which also include a series of non-market effects, one of which is increased health productivity.

Of course, these sort of arguments only hold if the correlation can be interpreted as a causal effect of education on health but, as Grossman (1975) has pointed out, there are at least two other possible explanations which have to be considered. The first is that some causality runs from better health to more education. Clearly, the possibility of this reverse causality is only present from childhood to early adulthood when education is being received. It cannot account for some of the observed schooling-health correlation among adults who have finished formal education. The second possible explanation, however, which states that there is no causal relationship at all but that the correlation is generated by the influence of one or more 'third factors' (confounders) seems to be more important. Such third variables could be genetic factors (e.g. mental ability) or early childhood environmental factors (e.g. parents' schooling).

More recently a 'time preference' hypothesis has been added to the third factor explanations (Fuchs, 1982; Farrell and Fuchs, 1982). It states that individuals with a high time preference (i.e. those who value future costs and benefits much lower than current costs and benefits) are less willing to invest in both education and health because typically the benefits of these investments accrue much later than the opportunity costs involved due to forgone time and earnings. Whatever the third factor may be, the larger its effect on schooling and health, the larger the spurious correlation and therefore the larger the overestimation of the effect of education on health when the third variable is omitted from the analysis.¹

The objective of this second part of the thesis is to test and identify the direct effect of education on individual health status, and the indirect effects through possible intervening variables. From the previous discussion it has become clear that an appropriate empirical test of this relationship requires both a firm theoretical basis and an adequate data set and corresponding methodology. In order to justify our own empirical research strategy adopted in the next two chapters, we will first review the theoretical and empirical issues raised in previous economic studies on this subject.

5.2. The effect of education on health demand and production

When we reviewed the economic theory with respect to health behaviour in Chapter 2 of this dissertation, we pointed out that there exist mainly two competing hypotheses regarding the effect of education. The first one is due to Grossman (1972a). He uses the standard household production assumption that education operates as an efficiency-increasing factor in both market and non-market production. He specified the following Cobb-Douglas gross investment in health function: (Grossman, 1972)

$$\ln I_t = \alpha \ln M_t + (1-\alpha) \ln T_t^n + \beta E \quad (5.1)$$

This means that health increments (I_t) are produced by combining medical care M_t (or market goods in general) and own time input T_t^n . The returns to scale are constant if $0 < \alpha < 1$, and education E increases non-market productivity if $\beta > 0$.² However, this basic hypothesis about the schooling effect is usually tested (with cross-section data on individuals) by estimating a reduced-form demand for health equation in which education enters with a coefficient equal to βe in the pure investment model, or to βy in the pure consumption model version. For the present analysis it suffices to know that, under reasonable assumptions about the marginal utility of healthy time and the marginal efficiency of health capital in producing healthy time, both e and y can be expected to lie between zero and one (see Wagstaff, 1986, pp 199-201). But even then, it means that a test of the positive education effect ($\beta > 0$) is conditional upon the assumptions about e and y and that any interpretation of this effect is also dependent on these assumptions.

An alternative assumption about the influence of education was made by Muurinen (1982). She does not adopt the household production framework but instead invokes the concept of capital use intensity to make the depreciation rate of health δ a negative function of education (see equation 2.16 in Chapter 2). Although this may seem "an intuitively more appealing device for handling environmental effects on health" (Muurinen, 1982, p. 9), Wagstaff (1986) shows that the difference is merely a matter of accounting convention if the depreciation rate function is assumed to be of the form (see 2.18)

$$\ln \delta_t = \ln \delta_0 + \beta_1 t + \beta_2 X \quad (5.2)$$

This implies that the initial depreciation rate δ_0 is augmented at an exponentially increasing rate as the individual ages ($\beta_1 > 0$ is also assumed by Grossman) but also increases with the environmental variables included in the vector X . If education is one of the X variables, then its coefficient in the demand for health function becomes $-\beta_2 e$ in the pure investment model and $-\beta_2 y$ in the pure consumption model (Wagstaff, 1986, p. 200). Since this implies that $\beta_2 = -\beta$, the two approaches cannot be distinguished at an empirical level by estimating demand for health functions.

A second problem with estimating demand for health functions is that, in order to avoid the use of non-linear estimation methods, two sub-models have to be estimated which sometimes give rise to very different results. In Chapter 2 we already mentioned Wagstaff's (1986) finding that the coefficient of education in the structural demand for health equation was significantly positive in the pure investment (PI) model but switched to significantly negative when the marginal utility of initial wealth $\lambda(0)$ was controlled for in the pure consumption (PC) model. This division into submodels is somewhat unsatisfactory because clearly health is demanded both for the direct utility it generates ('feeling fit') and for its production benefits ('making someone capable of undertaking market and non-market production activities'). Developments in non-linear estimation techniques may make this problem no longer insurmountable.

Thirdly, an important, though sometimes neglected, issue concerns the nature of the disturbance term in the demand for health equations derived by Grossman (1972a). It equals $-\epsilon \ln \delta_0$ in the PI model and $-\gamma \ln \delta_0$ in the PC model where ϵ and γ are positive constants and δ_0 is the depreciation rate in the initial period. In both cases, it is treated as an error term with zero mean and constant variance. However, it is highly unlikely that variations in δ_0 , the rate of health depreciation early in the life-cycle (e.g. at age sixteen) when individuals begin to make decisions themselves, is randomly distributed and uncorrelated with other explanatory variables in the demand equation. In particular, it is conceivable that it depends on genetic inheritance, early childhood environment and perhaps other factors which also influence intelligence or learning ability and thereby the level of schooling (Grossman, 1975). In other words, the omitted variables influencing δ_0 may be correlated with schooling and the correlated with schooling and the negative covariance between error term and explanatory variable may cause an upward bias in the estimated coefficient of schooling in the demand for health equation. This is essentially the argument presented in the introduction to this chapter about a third factor possibly causing a spurious correlation. In the next section we will discuss the possible remedies to this problem of omitted variable bias.

Furthermore, testing of the demand for health function approach becomes difficult when the 'observed' **actual** stocks of health and the unobserved **desired** stocks differ because adjustment does not take place instantaneously (Wagstaff, 1986; Grossman, 1975). The derived demand hypothesis, i.e. that more medical care and other health inputs are demanded the higher the desired health stock, does not necessarily hold for the actual current health stock. On the contrary, without exception all investigations of medical care utilization find a strongly negative relationship with observed health status. This suggests that either the derived demand hypothesis, and thus the essence of the Grossman model, has to be rejected or that costs of adjustment from actual to desired stock have to be incorporated into

the basic model set up.

Finally, important arguments in the demand for health function are the prices of the (endogenous) health inputs. In individual cross-section studies estimation of these price effects proved to cause problems because: (a) medical care prices typically show little variation between individuals and (b) time prices are difficult to impute to those not earning an hourly wage rate. Researchers have solved (circumvented?) these problems by using variation in regional prices or (sometimes endogenous) insurance coverage, and by restricting the sample to labour force participants, but both these solutions and their results are not very satisfactory. Medical care prices are often found not to have a significant influence on health demand and the limitation of the sample to labour force participants excludes a number of important groups of medical care users like the aged, disabled, housewives and children. Evans (1976) therefore argues that: "The result is to make any generalizations from these models to aggregate health care utilization very suspect" (Evans, 1976, p. 536). Although the basic idea that health increments are produced by combining medical care use and own time input may be sound, the difficulties to operationalize the relevant input prices for people with extensive insurance coverage for medical expenditures and sick leave, or for people not in the labour force, makes the testing of these assumptions through estimation of demand curves for health problematical.

The foregoing discussion serves to illustrate the difficulties involved in testing the predictions of the demand for health model empirically. However, for our purposes, an alternative is available which avoids most of the problems mentioned earlier. Since we are not so much interested in testing all of the Grossman model predictions but only want to obtain a good estimate of the education effect on health, we could use a more direct approach by focussing on production functions rather than demand functions for health. Of course, estimation of these technical relationships does not tell us much about rational economic behaviour. It only provides us with information about the production technology without assuming anything on choice behaviour given relative prices and limited resources. This kind of information is sometimes supplied by engineers in the case of manufacturing industries and is comparable to the outcome of epidemiological studies in the case of health. Nevertheless, because of the somewhat different orientation of economists interested in returns to scale and substitution possibilities at a fairly aggregate level, they have very often resorted to estimating production functions themselves, also for health (see e.g. Grossman (1972a), Newhouse and Friedlander (1980), Auster et al. (1969), Leigh (1983) and Kemna (1984)).

Ideally, what one then would like to estimate are gross investment equations like (5.1). However, it is obvious that these cannot be

estimated directly because the health increments I_t produced in each period are not directly observable. A solution to this problem, suggested by Grossman (1972a), is to combine the investment function (5.1), the depreciation rate function (5.2) and the definition of health change over time (equation 2.3 in Chapter 2) to obtain the following function: (for a derivation see Grossman, 1972a, p. 43)³

$$\ln H_t = \alpha \ln M_t + (1-\alpha) \ln T_t^h + \beta E - \beta_1 t - \beta_2 X - \ln \delta_0 \quad (5.3)$$

or, in general form

$$H_t = f(M_t, T_t^h, E, t, X, \delta_0) \quad (5.4)$$

Although the dependent variable H_t is now directly observable, a few observability problems remain with the arguments of this function. The investment inputs medical care M_t and time spent investing in health T_t^h are difficult to measure, and if they can, they are likely to be positively correlated with the disturbance term when δ_0 cannot be measured. The downward bias caused by this correlation explains why some researchers (e.g. Grossman, 1975; Kenna, 1984) found a large and significant negative impact of their measure of medical care use M_t on health status H_t in OLS estimations using individual cross-section data. When a two stage least squares procedure was used to mitigate this bias, difficulties were encountered to calculate predicted values of the endogenous inputs: most (truly demand) instrument variables used in the first stage turned out to be poor predictors and the second stage results were very sensitive to the particular specifications employed.

When data are available about past health status, a much simpler solution to this problem of omitted variable bias can be used. From the definition of net health change per period, one can immediately derive an estimable equation as follows:

$$H_t = H_t - H_{t-1} = I_t - \delta H_{t-1} \quad (5.5)$$

$$\text{or } H_t = (1-\delta) H_{t-1} + I_t \quad (5.6)$$

where I_t and δ can be substituted by their respective functions. Moreover, when the net change in the health stock H_t can be observed directly, it becomes possible to discriminate between the two competing hypotheses about the influence of education. Recall that the use-related depreciation hypothesis states that education reduces the use intensity and thereby the depreciation rate δ , whereas the productivity-increasing hypothesis postulates a direct effect on health investment I_t . The functional forms for δ and I_t given in equations (5.1) and (5.2) could directly be substituted in (5.5), but this would make the equation non-linear and therefore difficult to estimate. Alternatives will be suggested in the next section.

In this section we have contrasted (a) the two competing hypotheses regarding the effects of education (capital) in demand for health models, and (b) the two types of functions which can in principle be estimated to quantify the health effects of education. It was argued

that a number of theoretical and empirical problems could be avoided by refraining from estimating demand for health functions and instead focusing on production functions. When longitudinal data on health status are available, it does not only become possible to substantially reduce the problem of omitted variable bias but also to distinguish empirically the use-related depreciation versus efficiency-improving effect of education. In the next section, we will review the design and results of some empirical investigations and derive some implications for our own approach.

5.3. Methodological issues: a survey of previous research

Previous studies of the (causal) relationship between an individual's level of schooling and level of health can be subdivided according to a number of criteria. Theoretical criteria include whether or not they were based on a formal model of economic behaviour and whether they specified behavioural (i.e. demand) or technical (i.e. production) relationships. Using empirical criteria one could distinguish between cross-section and longitudinal data investigations, between studies which did or did not control for confounding 'third variables' or which did or did not examine intervening variables or reciprocal causation.

Figure 5.1. Various possible causal relationships between schooling and health

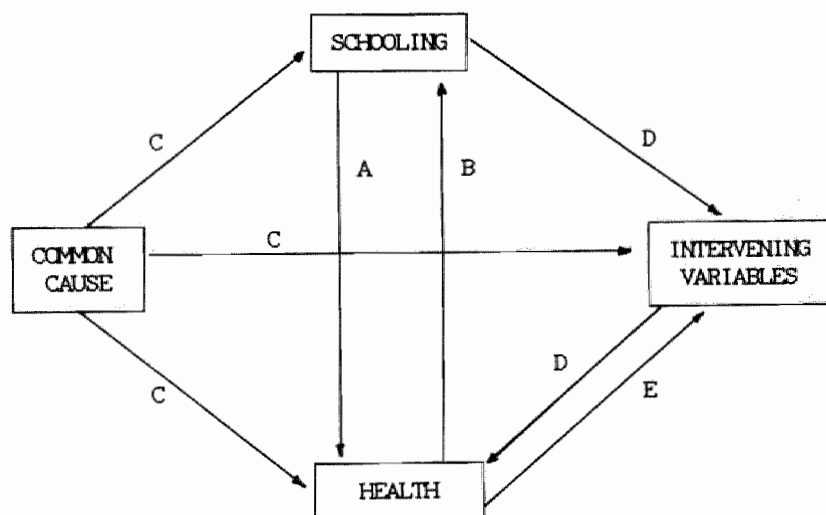


Figure 5.1 provides an overview of the various possible relations that could be considered. We are mainly interested in identifying the total effect of schooling on health which is the sum of the direct effect (A) and the indirect effects (D) via intervening variables like choice behaviour.

5.3.1. Reverse causality (B)

In order to disentangle the effects A and B, it is clear that longitudinal data are required about pre- and post-education health level. The temporal ordering of events then permits to investigate the causal effects in both ways. This approach is followed by Grossman (1975) who estimated a recursive health-schooling model using data on a sample of 3534 highly educated men who were married and in the labour force in 1969. A retrospective question was used to measure whether their health status was excellent, good, fair or poor when they were attending high school. With this past health measure and third variables (e.g. parents' schooling and ability measures) held constant, schooling is found to have a positive and statistically significant effect on the current subjective evaluation of health. However, also past health has a significant positive effect on years of formal schooling completed. The elasticity of current health with respect to schooling is .18 and dominates the reverse elasticity of schooling with respect to past health which only equals .03 (Grossman, 1975, p. 186).

Reciprocal causation was also tested by Shakotko, Edwards and Grossman (1981) between health and cognitive development of adolescents. The longitudinal data they use stem from Cycles II and III of the Health Examination Survey of the US National Center for Health Statistics. Their panel consists of 1434 white adolescents who were between 6 and 11 years in the Cycle II period (1963-1965) and 12 to 17 years in the Cycle III period (1966-1970). Cognitive development is measured by an IQ measure derived from an intelligence test for children and a school achievement measure derived from an arithmetic and reading test. Six measures of childhood and adolescent health were used, including indicators of oral health, obesity, blood pressure, presence of significant abnormalities, parent's assessment of overall health and school absence for health reasons. They estimated eight linear equations to explain the adulthood values of all eight measures by their childhood values and a number of family background variables.⁴ They found that the cross-lagged effects of cognitive development measures on health indicators are larger than the reverse effects. This is again interpreted as an indication that causality runs more strongly from cognitive development to health than vice versa.

5.3.2. Third factors (C)

For our purposes, the problem of reciprocal causation can fairly simply be avoided by restricting the analysis to adults who have completed their formal education. Health during adulthood does not influence educational achievement during adolescence. However, the danger of spurious correlation caused by one or more third factors (C in Figure 5.1) remains. One approach to solve this problem in cross-sectional studies has been, of course, to include measures or proxies for possible third variables like parental characteristics (e.g. Shakotko, et al., 1981; Edwards and Grossman, 1983) or mental ability test scores (e.g. Grossman, 1975). But some factors, like genetic health endowments, are inherently unobservable and therefore bound to be omitted in any specification of a health equation. Nevertheless, if such factors can be assumed to operate once and for all and if the main aim is not to distinguish environmental from genetic effects, a straightforward solution to this measurement problem is to use longitudinal data. Observations at two points in time on the dependent variable may be sufficient to avoid the bias which is caused by omitting the relevant but unobservable variable.

Consider the following simplified two-period model which is similar to the one described by Shakotko et al. (1981):

$$H_t = a_1 H_f + b_1 X + \epsilon_1 \quad (5.4)$$

$$H_{t+1} = a_2 H_f + b_2 X + c_2 H_t + \epsilon_2 \quad (5.5)$$

where H_t is health status in period t and H_{t+1} in period $t+1$, H_f represents an individual-specific but time-invariant third factor (genetic or environmental); X is a time-invariant background variable and ϵ_1 and ϵ_2 are disturbance terms. If H_f cannot be measured, estimation of b_1 in equation (5.4) leads to omitted variable bias when H_f and X are correlated.⁵ However, b_2 can be estimated without such bias if it can be assumed that a_2 is zero, i.e. that the impact of the third factor operates once and for all and is fully embodied in past values of the dependent variable.⁶ Under the latter, fairly restrictive assumption, Shakotko et al. (1981) estimate the - what they call - 'dynamic' effects of such family background variables as mother's and father's education and family income. If the lagged level of health is controlled for, the estimated b_2 coefficients of the X variables represent **short-run** effects free of genetic bias. These estimates are to be distinguished from the b_1 coefficients obtained from equations without lagged health status which then measure the 'cumulative', long-run impacts of such background variables.

Shakotko et al. (1981) estimated all health (and cognitive development) equations with and without lagged health included. Not surprisingly, they found that the magnitudes of the 'dynamic' effects were much smaller than the 'cumulative' effects. Only the level of mother's schooling showed some significant beneficial influence on adolescents' health status when childhood health status was controlled

for, and this in only three of the six health equations.

Distinguishing between so-called dynamic and cumulative effects is basically the same approach as the one adopted by Taubman and Rosen (1982) who estimated what they call 'health deterioration functions' of the following form:

$$H_{t+1} = H_t + \Delta H_t = f(H_t, X, t) \quad (5.6)$$

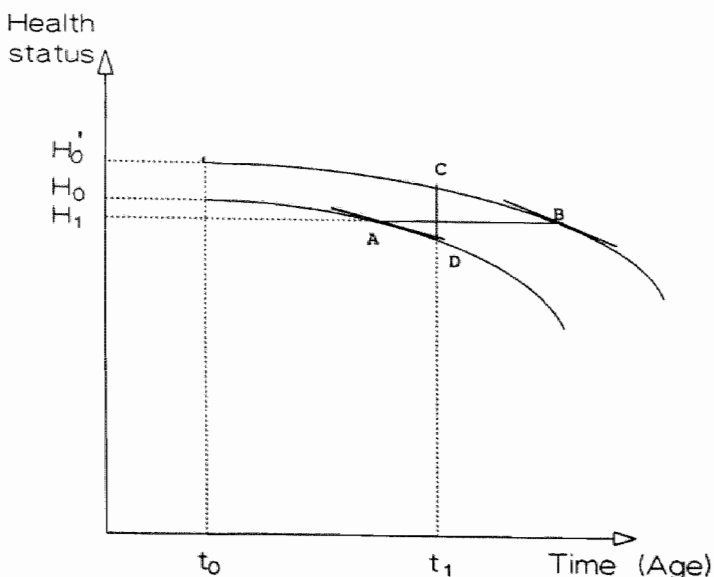
where the variables are the same as in equation (5.5) but the functional form is not necessarily linear.

Given an initial condition H_0 , difference equation (5.6) can in principle be solved to obtain the following equation for H_t

$$H_t = g(H_0, X, t) \quad (5.7)$$

When equation (5.7) is estimated in a cross-section and H_0 is treated as an unobserved random variable, the - by now familiar - problem of omitted variable bias occurs. Taubman and Rosen (1982) therefore suggest to estimate (5.6) instead. Of course, the estimated effects of X 's on health changes over time are not directly comparable to their effects on health levels (cfr. the distinction between dynamic and cumulative impacts). The difference can be illustrated by means of Figure 5.2 which is similar to the one presented in Taubman and Rosen (1982, p. 124).

Figure 5.2. Two hypothetical health profiles over the life cycle



Suppose the life-time profiles of the health of two individuals with different levels of education (E_1 and E_2) can be represented by the lines drawn in Figure 5.2. If the difference in initial health status ($H'_0 - H_0$) at point t_0 is unobserved, a cross-section estimate of equation (5.7) measures the distance between points R and S at some point t_1 . Equation (5.6), however, estimates the slope difference at points P and Q, given health status H_1 . This is zero if the profiles are parallel.

Taubman and Rosen (1982) used data from three waves of a panel survey to examine health changes over time. The Retirement History Survey was held among a random sample of some 11,000 men and women between the ages of 58 and 63 in 1969, 1971 and 1973. Health was measured as self-assessed health status compared to others of the same age with three answer possibilities (better, same or worse). Since it is difficult to calculate first differences of these health values without an arbitrary scaling method, the authors instead used a loglinear model to analyse effects and interactions of explanatory variables. Conditional on previous health level, they examined models explaining the proportions in these three level categories⁷ and in three categories expressing health change since the previous interview: improving, same, worsening. The authors conclude that 'health at one time and its change over time are strongly related to education and marital status, even when controlling for family income, use of medical resources, and previous health' (p. 140). However, not all of these results can easily be verified. The article does not report the size of the sample that was analyzed. It must certainly have been smaller than 11,000 since apparently only the data on men were used, each survey contains some missing data and panels also suffer from attrition. But even if the sample is close to 10,000, the cell frequencies soon become very small when a few interactions among variables are considered.⁸ No likelihood ratio and Chi-square tests are reported but it is stated that "nearly all the variables discussed in the text are statistically significant" (Taubman and Rosen, 1982, p. 121).

5.3.3. Intervening variables (D)

Finally, a number of researchers have looked at the possible intervening variables through which schooling may affect health. Unfortunately, most of those studies suffer from the deficiencies of cross-section data. Leigh (1983) explicitly attempts to test for direct and indirect effects of education on health (measured on a 7-point self-assessed individual health scale) through choice of occupation and healthy habits. He estimates a recursive model on US cross-section data taken from the University of Michigan's Quality of Employment Surveys for 1973 and 1977. The indirect effects of schooling through less smoking, more exercising and choosing less hazardous occupations far outweigh any direct effects. However, the results of

this study have to be interpreted very cautiously. Because no control for initial health status was made, and because the results differ very much between the two years, it is dangerous to make causal inferences like the author does. Moreover, due to a lack of theoretical framework, it is unclear whether the health equation is to be interpreted as a production or demand function.

Two studies which did make explicit the underlying hypotheses and did attempt to control for initial health are Grossman (1985) and Kemna (1985). Grossman (1975) decomposes the total schooling effect on health into a direct (technical) and three indirect (allocative) effects which are supposed to represent increased ability to select a better health input mix. The three endogenous inputs considered are: wife's schooling, job satisfaction and the difference between actual and ideal weight (as a proxy for diet selection). The estimated indirect effects through these three intervening variables are all strongly significant but represent only about 37% of the total schooling effect (controlling for initial health).

Kemna (1985) examines three possible intervening factors in the schooling-health relationship: hazardous occupation, demand for doctor visits and weight. Doctor visits is deleted from the decomposition analysis because it is found to have a strongly negative impact on health, according to the author due to unobserved differences in initial health not adequately controlled for by the proxy used (presence of chronic condition). The indirect effects of obesity and job hazards account for only a minor share, respectively 6 and 10% of the total schooling effect on health.

It is almost impossible to make any generalization from the above studies due to the large differences in samples analyzed, specifications used and results obtained. The main difficulty clearly seems to be the lack of longitudinal data in order to test for causality. A somewhat different research strategy was adopted by Farrell and Fuchs (1982). They examined the smoking behaviour of different cohorts of men and women ($n = 1183$) before and after they had completed their formal schooling. Their principal finding is that the negative relation between completed level of schooling and smoking is generally as strong at age 17 as it is at age 24 for all cohorts. They, therefore, reject the hypothesis that schooling differences are causal to smoking differences in favour of a 'third variable' explanation of the association. However, a second important finding is that the relationship does appear to be health-motivated because it developed only after the spread of public information about adverse health effects of smoking. The authors therefore conclude that cigarette smoking is undoubtedly an important intervening variable in the schooling-health relationship but that the link between education level and smoking is not causal. They are even able to reject the hypothesis that social class differences could be the 'omitted third variable' since father's schooling does not show any statistically

significant effect at all.

In sum, it appears that cohort analysis is one of the more promising research strategies to be adopted if causal inferences are to be drawn from empirical investigations of the schooling-health relationship.

5.4. Conclusions

In this chapter we discussed the theoretical and empirical problems associated with the analysis of the schooling-health relationship. It was argued that education effects on the health of the general adult population could better be examined by means of production functions than demand functions. The advantage is that one does not have to deal with a number of specific problems inherent to the demand for health approach. On the other hand, most of the ingredients of a truly economic approach (e.g. rational behaviour under constraints) are deleted when an essentially technical relationship like a production function is estimated. For our purposes, and with the available data in mind, it was judged that the advantages of the production function approach outweigh the disadvantages.

On the methodological side, we identified mainly four issues: (a) omitted variable bias, (b) measurement error, (c) direct versus indirect effects and (d) non-linearity and interaction. With respect to the problem of omitted variable bias, it was shown that longitudinal data might provide an acceptable solution if the objective is not so much to identify 'omitted third factors' but rather to control for possible spurious correlation. Measurement problems are inherent to any investigation involving the multi-dimensional and hardly quantifiable concept of health. None of the studies reviewed attempted to combine several health measures into one index, if more than one indicator of health was available at all. All studies used years of schooling as the variable to quantify attained level of education.

Studies which explored the linkage between schooling and health through intervening factors gave some support to the hypothesis that a lot of the effect works indirectly through allocation decisions. But the problem with these studies is that probably the total effect is substantially overestimated as a consequence of omitted variable bias (issue (a)).

Finally, only a few studies (e.g. Taubman and Rosen (1982), Kenna (1985)) allowed for the possibility of non-linear and/or non-additive effects of education. In the next two chapters, we will present two attempts to estimate education effects in health production. No dataset was available which allowed us to tackle all of these methodological problems simultaneously. However, each of the two

datasets had different strengths and weaknesses compared to the other so that none of them could be discarded for an empirical analysis a priori. We, therefore, analyzed both and placed different emphasis on certain aspects.

Notes

1. Note that this problem is very similar to the one that was faced by researchers investigating the ability bias in the income-schooling relationship (e.g. Griliches and Mason, 1972; Chamberlain, 1977).
2. E is not necessarily restricted to be one variable (own education) but can, according to the household production view, be extended to a vector including also the levels of education of other household members (see e.g. Wagstaff, 1986).
3. Grossman (1972) did not include environmental X variables in the depreciation rate function. Note that if education is also included among the X variables to allow for a depreciation effect, its coefficient in equation (5.3) reduce to $(\beta - \beta^E_2)$ where β is expected to be positive and β^E_2 negative. This approach is therefore unable to discriminate between the two rival hypotheses.
4. Their interpretation of these (reduced form) equations as either demand or production functions is unclear, as it was also in Grossman (1975). Both input factor prices, which belong in a demand function, and quantitative measures of input factors, which belong in a health production function, are absent.
5. For a formal derivation of the magnitude of such an omitted variable bias, see e.g. Maddala (1979), p. 305.
6. This can be shown as follows: Rewrite equation (5.4) to obtain

$$H_t = \frac{1}{a_1} H_{t-1} - \frac{b_1}{a_1} X + \frac{e_1}{a_1}$$

Then substituting H_t in equation (5.5) gives:

$$H_{t+1} = \left| \frac{a_2}{a_1} + c_2 \right| H_t + b_2 \left| 1 - \frac{a_2 b_1}{a_1 b_2} \right| X + \left| \epsilon_2 - \frac{a_2}{a_1} \epsilon_1 \right|$$

$$\text{or } H_{t+1} = \gamma H_t + \beta \cdot X + \epsilon'$$

If $a_2 = 0$, then $\gamma = c_2$, $\beta = b_2$ and $\epsilon' = \epsilon_2$ which gives again equation (5.5).

7. In 1971 and 1973 there was also a fourth category: deceased.
8. For instance, only about $10 \times .04 \times .35 = 0.14\%$ of this sample is divorced, has had 16 or more years of schooling and considered their health status worse than others in 1969.

Chapter 6. The marginal health productivity of schooling: A longitudinal latent variable model

"I agree with his (Fuchs') contention that the mechanisms via which schooling affects health have not been fully identified. I think that panel datasets, which allow one to control for 'unobservables' can be used to shed light on this and other issues" (Grossman, 1982, p. 2)

6.1. Introduction

In this chapter we will construct and estimate a longitudinal model which is similar to the models specified by Shakotko et al. (1981) and Taubman and Rosen (1982). It differs from these previous approaches in at least three respects. First, we will pay special attention to the health measurement problem by combining several indicators to construct one index of general health using the MIMIC-approach. Secondly, by specifying a structural equation model we will be able to estimate simultaneously the 'cumulative' and 'dynamic' effects of education on health (cfr. Shakotko et al., 1981) and to decompose the latter effect into direct and indirect effects. And thirdly, we do not restrict the analysis to one particular age category but develop a model which is suitable for the adult (Dutch) population.

6.2. Data

The data used in this chapter stem from a two-wave panel survey that was held by the University of Limburg for the Nederland Oké Foundation in 1981 and 1983. The total sample analyzed here consists of the respondents of two subsamples which completed the questionnaires in both years. A more detailed description of the stratified, multi-stage random sampling design can be found in Adriaanse et al. (1981). In 1981, the first subsample consisted of 1278 respondents drawn from the

general Dutch (active) population in the age-group 21 to 65. The second subsample contained 247 respondents drawn from the so-called non-active population under 65. The latter consists of the following categories: unemployed for at least six months, disabled, early retired or not working due to illness for at least three months already. In 1983, respectively 77% (n = 884) of the original active population sample and 75% (n = 153) of the non-active population sample completed the questionnaire (Halfens et al., 1984). After deletion of missing observations on some of the relevant variables for this study, 895 was the combined panel size retained for analysis. No correction was made for attrition or for the oversampling of non-labour force participants which makes this sample somewhat less healthy than the general Dutch population under 65. The interviews included a large number of questions on health and illness status, opinions, attitudes, beliefs and behaviour and on health care utilization. The variables which were selected and constructed for this study are described in table 6.1.

Table 6.1. Description of variables from the Nederland Oké Survey
(n = 895)

	mean	st.dev.
Health status variables		
SRHB1 = self-rated health status in 1981 (5 categories ranging from very good (5) to very bad (1))	3.915	.818
SRHB3 = self-rated health status in 1983	3.923	.754
CHRONB1 = a dummy (0/1) variable indicating the presence of at least one chronic condition or handicap in 1981	.284	.451
CHRONB3 = equivalent in 1983	.285	.452
CCMPLB1 = a likert sum score on a list of 21 complaints (0 to 21) (see also Appendix 6.A)	4.670	4.483
CCMPLB3 = equivalent in 1983	4.180	4.208
Background variables¹		
FEM = a (0/1) dummy variable for female sex	.458	.499
AGE = age in years	45.112	12.973
EDUCB3 = highest educational level attended (i.e. not completed) in 1983 (4 categories rang- ing from primary school (1) to university and non-university higher degree (4))	1.963	1.128
DEEDUC2 = dummy variable for lower secondary education (EDUCB3 = 2)	.240	.427
DEEDUC3 = dummy variable for higher secondary education (EDUCB3 = 3)	.105	.307
DEEDUC4 = dummy variable for university and non- university higher education (EDUCB3 = 4)	.171	.377

SINGLE	= one-person household (incl. unmarried, widowed and divorced)	.095	.293
JOB	= having a paid job (0/1)	.456	.498
PRIV	= private (or civil servant) health insurance coverage (0/1)	.328	.470
DISTGP	= distance in hundreds of meters to GP	14.668	15.331
ILL	= not working due to illness for more than three months (0/1)	.007	.082
SUNEM	= unemployed for less than six months (0/1)	.047	.212
LUNEM	= unemployed for more than six months (0/1)	.061	.240
RET	= retired (0/1)	.131	.337
DISAB	= receiving disability insurance benefits (0/1) (1983)	.009	.094
STUD	= student (0/1)	.011	.105
HOUSEH	= doing the household (0/1)	.278	.448
FINCC	= gross household income in 1981 in classes (8 categories ranging from less than 12.000 guilders (1) to more than 100.000 (8)) ²	4.879	1.716
HHSZ	= number of persons living from household budget in 1981	3.201	1.418

Behavioural variables¹

DOCVIS83	= number of doctor visits (GP and specialists) in 1983 (during last 12 months) (9 categories: 0, 1, 2, 3, 4, 5 to 7, 8 to 10, 11 to 15, 16 or more)	2.000	2.452
SMDKE	= smoker (0/1) in 1983	.515	.500
EXERC	= doing some physical exercise (0/1)	.463	.499
HLIFEST	= self-rated healthiness of life-style (6 categories ranging from unhealthy (1) to very healthy (6)) in 1983	4.522	.926

Notes

1. All variables refer to the 1983 values except when indicated otherwise.
2. If 1981 income was missing, 1983 was used instead if available.

6.3. Measuring health changes over time

6.3.1. Variability of health

The Nederland Oké Survey contained three indicator variables which can be used to measure the unobserved concept of individual permanent health stock: self-rated general health status (SRH), the presence of at least one chronic condition or handicap (CHRON) and a sum score on a list of 21 psycho-somatic complaints (COMPL). The Questionnaire for Research into Experienced Health status (GREH, or VÖEG in Dutch) was

developed to measure stress by Dirken (1967). A translation of the 21 questions is presented in Appendix 6.A. The list used here is a shortened version of the full questionnaire including 47 items.

Self-rated health status (SRH) is the most commonly used (subjective) measure of 'permanent' or 'overall' health status of individuals in all previous studies. CHRON is a more objective measure of physical health status, and COMPL was designed to measure the psycho-somatic impact of stress. Van Vliet and Van Praag (1987) have recently shown that these three variables are reliable indicators of the individual permanent health status index of a representative sample of the Dutch population aged 18 years and over ($n = 2153$) and strong predictors of health care utilization. However, for our purposes it is important to know whether these indicators are sufficiently sensitive to measure any health changes over a relatively short time period of two years.

Table 6.2. Cross-tabulations of health variable frequencies for 1981 and 1983

a. Self-rated health status

SRH81	SRH83					TOTAL
	very bad	bad	fair	good	very good	
very bad	1	1	1	-	2	5
bad	2	10	30	8	1	51
fair	2	8	119	66	9	204
good	-	5	57	378	98	538
very good	-	2	13	110	111	236
TOTAL	5	26	220	562	221	1034

b. Psychosomatic complaints

COMPL81	COMPL83					TOTAL
	0	1-5	6-10	11-15	16-21	
0	96	67	12	2	1	178
1-5	106	320	64	9	-	499
6-10	12	99	101	24	3	239
11-15	1	17	38	30	8	94
16-21	1	2	2	9	11	25
TOTAL	216	505	217	74	23	1035

c. Chronic condition or handicap

CHRON81	CHRON83		TOTAL
	yes	no	
yes	216	71	287
no	72	674	746
TOTAL	288	745	1033

In table 6.2 we have presented the cross-tabulations of these variables for 1981 and 1983. From these tables, transition probability matrices could be estimated if all cells had sufficient observations. However, it can be seen that the large majority of respondents is located on the diagonal, which means that their health status did not change, according to our measures, in the two-year period. For CHRON the diagonal contains 86.1% of the sample, for SRH this percentage is 60.5, and for COMPL only 54% even if within-class off-diagonal elements are not subtracted. Clearly, COMPL offers most of the dynamic variability but is also has got some disadvantages: the complaint items, of course, do not cover all possible health complaints but only a selection of rather vague a-specific physical complaints. Moreover, within these 21 items, some 'health problems areas' have disproportionately large representation because the questionnaire was originally designed to measure stress: no less than 5 items refer to digestive complaints, 4 items refer to respiratory problems, 6 items refer to general fatigue or mental depression, and 2 items only refer to pain (back and headache) complaints. Van der Zee (1982) identified these four factors in a varimax rotated solution of a factor analysis performed on the respondents of the 1977 Dutch Life Situation Survey ($n = 4147$). On the other hand, some health problems are hardly or not represented: an example is restriction of mobility, the first principal component identified by Wagstaff (1986) in the analysis of a similar list of health questions with Danish data. As a consequence, our COMPL sum score is particularly sensitive to some health problems (e.g. stomach complaints) and very insensitive to others. Nevertheless, we abstained from extracting principal components or factors because our explicit intention is to measure **general** health and not **specific** health dimensions. By combining the COMPL variable with SRH and CHRON, we hope to correct for measurement errors created by the arbitrary summation procedure.

Finally, it is worth looking at the means of the health indicators in both years presented in table 6.1. For an ageing panel, one would expect the deterioration of health to dominate the recovery, but the reverse seems to be true: the means of SRH and CHRON are virtually unchanged and the mean COMPL score in 1983 is even slightly lower than in 1981! The latter result may be a consequence of the fact that in 1983 the list of complaints had been extended to cover all 47 original items which might have made the respondents a little more reluctant to give the yes answer too often (see e.g. Halfens et al., 1984). However, not too much attention should be given to the comparison of means because this does not incorporate variations in the distributions around these means. More appropriate methods are considered in the next section.

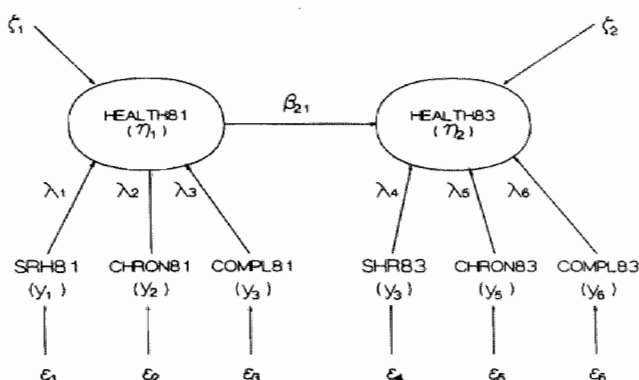
6.3.2. Stability of latent health status

In Chapter 4 we have described how the LISREL model can be used to

characterize latent, unobservable variables by observable indicators. In this chapter, we will use a similar measurement strategy but in addition exploit the possibility to examine the intertemporal covariance of measurement errors when using longitudinal data.

A two-wave three-indicator health measurement model can be represented conveniently by the following path-diagram.

Figure 6.1. Path diagram of a two-wave three-indicator model of health in 1981 and 1983



The LISREL notation (see e.g. Jöreskog and Sörbom, 1977) is used to characterize variables, coefficients and disturbances. By fixing λ_1 and λ_4 equal to one, the latent variables get the same unit of measurement as the self-rated health (SRH) variables. An alternative identification procedure is to fix the variance of the unobservable variables to one:

$$\Psi_{11} = E(\xi_1 \xi_1) = \Psi_{22} = E(\xi_2 \xi_2) = 1 \quad (6.2)$$

An advantage of the latter procedure is that it directly gives the standardized solution of the model which facilitates the testing of the psychometric properties of the measurement instruments (see Chapter 4, section 4.2). A disadvantage is that one cannot give a scale and direction to the latent variables in this way. We will report both parameterization results where appropriate.

We have first estimated two measurement models (I and II) separately for each year. Because the covariance matrix of the observed variables was used as the input matrix, the reported parameter estimates (λ 's, β 's and γ 's) can be interpreted as the usual regression coefficients. The results are given in the first two columns of table 6.3. The

Table 6.3. Parameter estimates for the sequence of measurement models (n = 895)

Parameters	I	II	III	IV	V	VI
λ_{y1}	1.00		1.00	1.00	1.00	1.00
λ_{y2}	-.390 (14.4)		-.479 (17.8)	-.430 (15.7)	-.401 (16.9)	-.404 (15.5)
λ_{y3}	-4.646 (15.3)		-5.902 (21.8)	-6.341 (22.3)	-4.833 (20.0)	-4.886 (17.0)
λ_{y4}		1.00	1.00	1.00	1.00	1.00
λ_{y5}		-.479 (13.9)	-.540 (16.8)	-.488 (15.0)	-.456 (15.9)	-.461 (14.5)
λ_{y6}		-5.441 (14.2)	-6.214 (20.3)	-6.671 (20.6)	-5.112 (18.5)	-5.177 (15.7)
ψ_{11}	.477 (11.8)		.358 (11.9)	.355 (11.8)	.458 (13.3)	.452 (12.0)
ψ_{22}		.323 (10.3)	.000 (.10)	.033 (4.0)	.077 (6.8)	.076 (6.6)
θ_{e1}	.192 (6.8)		.311 (17.1)	.314 (16.9)	.211 (10.9)	.217 (8.7)
θ_{e2}	.131 (17.6)		.121 (18.9)	.139 (19.5)	.129 (18.5)	.129 (10.1)
θ_{e3}	9.797 (13.2)		7.629 (15.2)	5.807 (11.8)	9.434 (15.5)	9.328 (13.7)
θ_{e4}		.245 (10.6)	.295 (17.8)	.297 (17.6)	.228 (13.1)	.233 (10.6)
θ_{e5}		.130 (16.9)	.124 (18.9)	.140 (19.4)	.132 (18.4)	.131 (18.0)
θ_{e6}		8.138 (11.5)	7.150 (15.4)	5.603 (12.3)	8.756 (15.7)	8.652 (14.1)
β_{21}			.875 (18.6)	.820 (17.9)	.758 (19.2)	.757 (19.2)
θ_{e52}				.045 (13.5)	.073 (12.8)	.072 (12.4)
θ_{e63}					4.849 (10.5)	4.754 (9.1)
θ_{e41}						.006 (.36)
χ^2	-	-	430.9	147.1	9.6	9.5
df	-	-	8	7	6	5
P	-	-	< .001	< .001	0.14	0.09

Note: The estimated coefficients can be interpreted as the usual regression coefficients. Figures in parentheses are absolute approximate t-values (coefficients divided by standard errors).

estimates of both the magnitude and statistical significance of the λ -coefficients and the variances of the ϵ 's (the θ_{ϵ} 's) prove to be very similar in both fits which indicates a high stability of the measurement model over time. The goodness of fit cannot be judged on the basis of a Chi-square test because the number of degrees of freedom is zero: the six model parameters are exactly identified by the six input covariances between the y-variables.

In order to be able to calculate reliability coefficients for the health indicators, we re-estimated the models with the correlation matrix as input in order to obtain the standardized λ^* -coefficients. The reliability coefficients, defined as $\rho = \lambda^2$, are presented in table 6.4. SRH has the highest reliability and CHRON clearly the lowest. This is as one could expect since CHRON measures the stock of health in a dichotomous manner which clearly creates a lot of measurement error.

In a next step, we have estimated the measurement model III which was presented in Figure 6.1. This is not a pure measurement model because it also contains one structural parameter (β_{21}) to measure the stability of health over time.¹ As can be seen from table 6.3, the overall health of the sample has remained fairly stable between 1981 and 1983 (according to the value of β_{21} in model III). The reliability coefficients (see table 6.4) of CHRON and COMPL have increased while the opposite is true for SRH. Although all estimated parameters are highly significant (most t-values have increased substantially compared to the separate model fit) the model is rejected by the Chi-square test ($\chi^2_{(8)} = 430.9$). This could mean that this model is too restrictive Jöreskog and Sörbom (1977) have pointed out that in this type of

Table 6.4. Reliability coefficients (ρ) of the health indicators in the measurement models

	MODEL			
	I	II	III	V
SRHB1	.713		.535	.684
CHRONB1	.357		.404	.365
COMPLB1	.513		.620	.531
SRHB3		.569	.481	.599
CHRONB3		.364	.391	.356
COMPLB3		.540	.596	.506

Note: The ρ -coefficients are defined as the squared standardized λ^* -coefficients.

panel models, when the same instruments (test or questionnaire) are used to measure the latent variable at both points in time, there may be a tendency for the errors in each variable to correlate over time because of memory or other re-test effects. Relaxing the zero restrictions on the intertemporal covariances between the ϵ 's of the same indicators can in that case contribute to a substantial improvement of the model fit.

The highest modification index for model III indeed was found for $\theta_{\epsilon 25} = \text{cov}(\epsilon_2 \epsilon_5)$. It equals 232 which indicates that the χ^2 -test could be reduced by at least 232 when the error terms of CHRONB1 (= y_2) and CHRONB3 (= y_5) are allowed to covary. This model modification does seem to make sense. It is reasonable to assume that the (rather large) measurement error made by measuring health by the dichotomous variable CHRON in 1981, is strongly correlated with the same measurement error in 1983.

The zero restriction on $\theta_{\epsilon 25}$ was relaxed to estimate model IV. The test statistic for this model is 147.1 for 7 degrees of freedom (see column IV in table 6.4). The reduction in the χ^2 statistic associated with the loss of one degree of freedom ($\chi^2_{(8)} - \chi^2_{(7)} = 430.9 - 147.1 = 283.3$) implies a statistically very significant improvement. Also the statistical significance of $\theta_{\epsilon 25} = .045$ (t-value is 13.5) is very high but the model still has to be rejected.

When in a next step, a non-zero covariance was also allowed for the errors of COMPL in both years, model V was obtained. This covariance θ_{ϵ} is estimated to be 4.849 and highly significant (t-value is 10.5). Moreover, the χ^2 -test statistic has now dropped to 9.6 which means an acceptable fit with 6 degrees of freedom and at a 10% confidence level ($P = 0.14$). When also correlated errors of SRH were allowed for in a next model VI, neither the estimated covariance, nor the reduction in χ^2 were significant. Apparently, only the measurement errors in CHRON and GREH are correlated over time.

We therefore choose model V as the measurement model that best describes the evolution of the individual general health status over time. The estimated λ 's and θ_{ϵ} 's are remarkably close to the original (separate) estimates in model I and II. The same is true for the reliability coefficients given in table 6.4. Note, however, that there has been a significant reduction in the magnitude of the stability coefficient β_{21} from model III to V. This implies that a substantial part of the stability measured in model III was due to the correlated errors of two of the indicators over time. The standardized β (not reported in table 6.4), which can be interpreted as the (corrected) correlation between the true health status of the sample individuals in 1981 and 1983, has similarly decreased from 1.00 in model III to .879 in model V. This is still, however, much higher than the intertemporal correlations between the same indicator variables which are .558, .679 and .713 for respectively SRH, CHRON and COMPL. We can

conclude that the correlation between the unobserved health status of individuals over time is higher than the same correlation between its indicators. Also this result seems to make sense and gives some confidence in the measurement procedure. In section 6.4 we will examine what happens to the stability parameter and the standardized correlation between the η 's when the latter are assumed to be determined by a number of background variables.

6.4. A structural equation model of health production

Rather than specify, estimate and test health production functions directly, we will proceed stepwise in this section. In the first step, we will explore whether the strength of the uncontrolled schooling-health association is as strong in our dataset as it has been found by others. Secondly, the dynamic and partial relationship is investigated by controlling for previous health and other determinants. In the third section, some health inputs into the production function are endogenized in order to test for direct and indirect effects through intervening variables. This stepwise procedure provides more insight into the various relationships considered and the separate contributions of explanatory variables.

6.4.1. The schooling-health correlation

The schooling level was measured in this survey by the highest of four possible levels of schooling that was followed (not necessarily completed) by the sample respondents in 1983 (EDUC83). Only about 1% of the sample was still receiving formal education at the time of the interview (see table 6.1). Because the lower age limit was 21, presumably most of these were receiving higher (university or non-university) education. table 6.5 presents the 1983 health indicator means by education level. No clear schooling gradient emerges from this table. All levels above primary school do show significantly higher levels of good health but between these three higher levels the differences are not always so clear and the patterns deviate depending on the indicator used. A peculiar finding is that the higher educated, though showing better health on fairly objective measures (CHRON and COMPL), are more critical of their own health status: the subjective rating (SRH) is not higher than that of the two previous levels of education. On the whole, the (uncontrolled) association seems weaker than what is usually found in American studies of this kind. The relation does not seem to be linear and the largest discrepancy is between those who did and those who did not receive more than primary education. Therefore, we will use three dummy variables (DEDUC2, DEDUC3 and DEDUC4) in what follows to measure level of schooling on an ordinal or ranking scale. The regression coefficients can then be interpreted as the health gain due to schooling above the primary

level.

Table 6.5. Health status by level of education (n = 1033)

Highest level of education attended	SRHB3 (mean)	CHRON83 (mean)	COMPL83 (% < 6)	Total n
Primary education and Vocational training	3.79	.326	69.3	512
Lower secondary education	4.07	.244	80.8	238
Higher secondary education	4.10	.255	79.1	110
Higher education	4.09	.199	83.1	171

6.4.2. Cumulative and dynamic effects

In this section, we gradually expand the bivariate relationship between schooling level and latent health, as it was constructed in section 6.3.2. We explained in section 5.2 that ideally one would like to estimate an explicit, linearized functional form of the production function for health like, for example, equation (5.3), which is basically Grossman's specification. In implicit form, this equation could be written as follows:

$$H_t = f(E, t, X, I_t, \delta_0) \quad (6.3)$$

where H_t = health status at the end of period t

E = education level

t = time (age in cross-section studies)

X = a vector of environmental variables affecting the depreciation rate δ

I_t = a vector of health investment inputs used in period t

δ_0 = the initial health depreciation rate

In order to illustrate the effects of not controlling for δ_0 we will estimate the following two-equation system simultaneously:

$$H_t = \beta_1 E + \beta_2 t + \beta_3 X \quad (6.4)$$

$$H_{t+1} = \beta_1' E + \beta_2' t + \beta_3' X + \beta_4' I_{t+1} + \beta_5' H_t \quad (6.5)$$

Here, for simplicity H_t and H_{t+1} refer to 1981 and 1983. Due to the lack of good measures of the health inputs I_t (own time and market goods), we did not attempt to estimate more sophisticated functional forms than a linear one. Restrictions or tests on the estimated coefficients (e.g. returns to scale tests) seem premature given the arbitrary scaling of the input proxies we used and the present state of knowledge regarding health production functions.

Table 6.6. Maximum likelihood estimation results of the LISREL two-equation structural model (n = 895)

	MODEL I			MODEL II			MODEL III			MODEL IV		
	HB1*	HB3*	HB1*	HB3*	HB1*	HB3*	HB1*	HB3*	HB1*	HB3*	HB1*	HB3*
HEALTH ^{HB1*}	-	.754 (18.9)	-	.752 (18.2)	-	.755 (18.0)	-	.734 (16.4)	-	.734 (16.4)	-	.734 (16.4)
DEDUC2	.387 (6.3)	.004 (.11)	.298 (5.1)	.002 (.06)	.219 (3.6)	.003 (.07)	.221 (4.0)	-.009 (.23)	.221 (4.0)	-.009 (.23)	.221 (4.0)	-.009 (.23)
DEDUC3	.308 (3.7)	.027 (.53)	.326 (4.9)	-.011 (.26)	.069 (.84)	.014 (.26)	.086 (1.1)	.019 (.37)	.086 (1.1)	.019 (.37)	.086 (1.1)	.019 (.37)
DEDUC4	.442 (6.4)	-.009 (.20)	.326 (4.9)	-.011 (.26)	.224 (3.2)	-.020 (.44)	.165 (2.4)	-.008 (.17)	.165 (2.4)	-.008 (.17)	.165 (2.4)	-.008 (.17)
AGE					-.017 (9.0)	-.001 (.49)	-.010 (5.0)	.001 (.42)	-.010 (5.0)	.001 (.42)	-.010 (5.0)	.001 (.42)
FEM			-.017 (9.3)	-.001 (.49)	.121 (2.6)	-.003 (.09)	-.054 (1.1)	-.026 (.80)	-.054 (1.1)	-.026 (.80)	-.054 (1.1)	-.026 (.80)
SINGLE							-.100 (.95)	.008 (.14)	-.100 (.95)	.008 (.14)	-.100 (.95)	.008 (.14)
RET							-.862 (11.4)	-.136 (2.3)	-.862 (11.4)	-.136 (2.3)	-.862 (11.4)	-.136 (2.3)
ILL							-.315 (1.2)	-.299 (1.7)	-.315 (1.2)	-.299 (1.7)	-.315 (1.2)	-.299 (1.7)
DISAB							.250 (1.1)	-.167 (1.0)	.250 (1.1)	-.167 (1.0)	.250 (1.1)	-.167 (1.0)
SUNEM							-.199 (1.9)	.072 (1.0)	-.199 (1.9)	.072 (1.0)	-.199 (1.9)	.072 (1.0)
LUNEM							-.091 (.90)	.072 (1.0)	-.091 (.90)	.072 (1.0)	-.091 (.90)	.072 (1.0)
HHSZ							.028 (1.5)	.014 (1.1)	.028 (1.5)	.014 (1.1)	.028 (1.5)	.014 (1.1)
FINCC					.069 (4.6)	.005 (.48)	.027 (1.8)	.001 (.11)	.027 (1.8)	.001 (.11)	.027 (1.8)	.001 (.11)
R ²	.085	.772	.190	.774	.223	.774	.376	.791	.376	.791	.376	.791
x ²	30.4 (p = .034)		64.7		82.0		123.7		82.0		123.7	
df	18		22		30		58		30		58	

Note: see table 6.3

Starting from the most simple model, with only E and H_t as explanatory variables, we then added the other arguments of (6.3) in a stepwise manner. The results are shown in table 6.6. The measurement equations are not reported here because they differ only marginally from those presented in table 6.3. In Model I, only the education dummies are used. In the equation for $HEALTHB1^*$, all education levels show a significantly better health status than the lowest level, but the coefficients do not increase monotonically with higher levels. The ranking order is similar to the one we saw in table 6.5 for two of the indicators (CHRON and COMPL) separately. The largest differences (and the only significant ones) occur between the lowest and the three highest levels. However, all of these differences vanish when in the equation for $HEALTHB3^*$ previous health status two years earlier is controlled for! This implies that, although the total (cumulative) effect of education on health status level is substantial, its effect on health changes over a two year period is negligible, even when no other influences are taken into account. In terms of the lifecycles pictures we drew in the previous chapter, this finding means that the time paths of health status have different intercepts but identical slopes, i.e. they are parallel. But before we make a digression on the interpretation of this result, we will first describe what happens when other health determinants are held constant.

In Model II, we have added AGE which is an important correlate of education in cross-sections of individuals. It can be seen that indeed controlling for age (linearly) reduces the schooling coefficients and shows a strong negative impact on health in 1983. However, like education, it does not affect the rate of deterioration significantly in the second equation. Similar results are obtained in Model III, when also household income and female sex are controlled for. Both variables have a significant contribution to the explanation of $HEALTHB1^*$, but none at all to the explanation of $HEALTHB3^*$ when $HEALTHB1^*$ is held constant. The inclusion of FINCC in the health equations needs some justification. Normally, price and income variables are not amongst the arguments of production functions. However, one can interpret the household income in this equation as a proxy for unobserved health input (see Grossman, 1975, p. 175) or as a proxy for living conditions (e.g. housing, nutrition, life style) which is one of the environmental variables affecting the depreciation rate.

The positive effect of FEM on health contrasts with the findings of a negative impact in most other studies of adult samples (see e.g. Chapters 4 and 7 of this book, or Van Vliet and Van Praag, 1987). However, it is most likely that this is a result of the oversampling of, mainly male, non-labour force participants for health or other reasons. In Model IV, we will also attempt to control for this peculiarity of the dataset. But first, we investigate whether the insensitivity of the latent health index to other variables besides lagged health is mainly a result of combining the information contain-

ed in each indicator. Table 6.7 gives the results of three separate regressions for each health indicator on its 1981 values and the set of explanatory variables contained in Model III.

Table 6.7. Maximum likelihood regression results of model III for the three separate health indicators (n = 895)

	SRHB3	CHRONB3	COMPLB3
SRHB1	.468 (17.1)	-	-
CHRONB1	-	.645 (24.7)	-
COMPLB1	-	-	.656 (28.3)
DEDUC2	.003 (.05)	-.009 (.30)	-.372 (1.4)
DEDUC3	.053 (.71)	.060 (1.5)	-.454 (1.3)
DEDUC4	-.022 (.34)	-.023 (.68)	-.243 (.79)
AGE	-.005 (3.2)	.004 (3.9)	.004 (.48)
FEM	.053 (1.3)	.011 (.50)	-.073 (.36)
FINCC	.039 (2.9)	-.003 (.42)	-.047 (.71)
R ²	.331	.473	.512

Note: see table 6.3

Remember that SRH is an indicator of 'good health' whereas CHRON and COMPL are 'bad health' indicators. It appears that also these separate indicators remain fairly unaffected by the other independent variables as soon as the two-year lagged variable is included. Only age still has a significant influence on SRHB3 and CHRONB3, and FINCC on SRHB3 at the 5% level of significance. The proportions of explained variance of these observed scores are lower than the R-squared for HEALTH83* (the 'time' score η_2) in table 6.6, Model III. This is mainly a consequence of the fact that correcting for measurement errors also reduces the variability to be explained. We can conclude that, except for AGE, the indicators are similarly independent of the background variables as the health index when past health is controlled for.

Finally, the X vector was extended with a number of variables which indicate the household or employment status of the individuals in order to obtain Model IV. Although we treated these characteristics as exogenous, causality cannot be inferred from these results because people may be in one of the considered categories (e.g. unemployed) for health reasons.² We do not find significant health differences between single-person and other households. Some of the non-active categories do show substantially inferior health status than the omitted category, which includes students and those working in a paid job or at home. But the effects are only statistically significant for the category of retired people in the sample (about 13%, see table

6.1). These are (mainly male) individuals who retired before the pensioner's age of 65 or who retired since 1981. Not only is their health status much lower than average in 1981 but it has also deteriorated faster than average between 1981 and 1983. As can be seen from the reduction in the coefficient of AGE, RET mainly captures the effect of accelerating health depreciation with ageing. Not surprisingly, also the respondents who are ill for more than three months already report a considerably higher health deterioration. The non-significance of these effects may be due to the limited number of people in this category (only 0.7% of the sample). The same applies to the DISAB category of people receiving a disability insurance allowance (8 respondents). Also the effects of being unemployed for less (SUNEM) or more than 6 months (LUNEM) are not significant at the 5% level of significance. As expected, FEM does now have a negative coefficient but is not significantly different from zero.

Although the results of Model IV show that it is not so that our health construct is unable to detect any differences when past health is controlled for, on the whole, the background variables considered here do not contribute much to the explanation of the (limited) variation in latent health status changes during the two year period. Like Shakotko et al. (1981), we can conclude that the dynamic effects of background effects are much smaller than the cumulative impact, if existing at all. Schooling does not have a significant marginal effect on changes in health and its effect on levels of health does not show a clear increasing gradient with levels of education. In the next section, we abandon the rather 'passive' view of health production adopted in this section to explore the health impact of behavioural choices.

6.4.3. Direct and indirect effects

In order to complete the health production function (6.3) that was specified in the previous section, the health inputs chosen by the individuals have to be added to the equation for HEALTH83*. In the original Grossman (1972) model, the two input factors in household health production were market goods and own time spent investing in health. Medical care use or expenditures were mostly used in previous studies to measure health investment outlays. No attempt has been made to use a proxy for the time input, but other health inputs used in empirical studies include choice variables such as job hazards, job satisfaction, wives' schooling and the difference between actual and ideal weight, given height (see Grossman, 1975; or Kemna, 1985 for examples). Leigh (1983) also considered smoking and exercising behaviour. Except for Kemna's (1986) choice of a hazardous occupation function, none of these 'other input' demand functions were derived from fully specified theoretical models. In most cases, this would require an extension of the basic Grossman model to allow for joint production effects because clearly smoking cigarettes and physical

exercise not only affect health but also produce other desired characteristics such as smoking pleasure or recreation (see e.g. Grossman, 1972, p 74). Such an extension of the theoretical model will not be pursued here. Our only aim is to decompose the total (dynamic) effect of schooling into direct and indirect effects. However, we saw in the previous section that this total effect was very small and non-significant. This implies that either also the direct and indirect effects are very unimportant or they cancel out in the aggregate.

To examine these effects through intervening variables, we have replaced the income variable FINCC by two endogenous health inputs. The number of doctor visits reported by the respondent in the last 12 months before the second interview (DOCVIS83) is used to measure the medical care input. The respondent's judgement regarding the healthiness of his/her life-style (ranging from 1 = unhealthy to 6 = very healthy) will be employed as a proxy for the time input (HLIFEST). The latter choice rests on two assumptions. First, that people reporting to have a relatively healthy life-style do on average have healthier habits than those who do not. And secondly, that a healthier life-style also implies that more time is spent on health preserving or health restoring activities.³ The demand functions for these two inputs comprise a similar set of predetermined variables as the health production function plus the income variable and, in the case of DOCVIS, a dummy for privately insured (PRIV). In contrast to those covered under the public sick fund scheme in the Netherlands, the privately insured often choose not to insure for GP services or take high deductibles (see e.g. Van de Ven and Van Praag, 1983). Consequently, the money price of a doctor visit is very often not zero for this category. The distance from the respondent's house to his GP (DISTGP) was added as a proxy for the (travel) time price of a GP visit.

The estimation results for the structural model with endogenous health inputs are presented in table 6.8. The two household characteristics (SINGLE and HHSZ) and the two unemployment variables (SUNEM and LUNEM), which did not show significant effects in Model IV, have been deleted from the X vector to restrict the total number of explanatory variables.

First of all, it can be seen that, as expected, HLIFEST and DOCVIS are strongly associated with initial health (HB1*): the better the 1981 health status, the better the life-style and the fewer doctor visits in 1983. More surprising are the effects of education: negative on life-style and positive on doctor visits but with only one dummy significant in each equation. Again, it is not impossible that the better educated are more critical in their judgement of their life-style. Women and older persons report healthier life-styles, which in this context could be interpreted as a result of lower average time prices, ceteris paribus. A similar interpretation holds for the negative effect of FINCC if this acts also as a proxy for the wage

Table 6.8. The structural equation model with endogenous health inputs

	HB1*	HB3*	HLIFEST	DOCVIS83
HB1*	-	.543 (13.6)	.411 (6.2)	-1.998 (11.8)
HB3*	-	-	-	-
HLIFEST	-	.054 (3.6)	-	-
DOCVIS	-	-.057 (8.0)	-	-
DEEDUC2	.257 (4.8)	.035 (1.0)	-.066 (.83)	.494 (2.6)
DEEDUC3	.128 (1.8)	.029 (.65)	-.105 (.97)	.267 (1.0)
DEEDUC4	.215 (3.5)	-.003 (.08)	-.210 (2.2)	.185 (.76)
AGE	-.011 (6.5)	-.001 (1.1)	.014 (5.3)	-.003 (.47)
FEM	-.042 (.95)	-.008 (.30)	.190 (3.0)	.578 (3.8)
RET	-.856 (12.1)	-.107 (2.2)	.000 (.001)	.398 (1.5)
ILL	-.320 (1.2)	-.228 (1.4)	.306 (.84)	1.389 (1.6)
DISAB	.103 (.45)	-.162 (1.2)	-.011 (.03)	.316 (.40)
FINCC	-	-	-.031 (1.6)	.015 (.29)
DPRIV	-	-	-	-.429 (2.4)
DISTGP	-	-	-	-.001 (.12)
R ²	.358	.841	.101	.319

Note: see table 6.3

rate and thereby for the opportunity cost of time. With respect to the demand for physician visits, females, retired and ill persons have higher and privately insured lower demands. Physician proximity (DISTGP) does not have an influence on visit frequency.

In the HEALTHB3* equation both input factors show a significant influence on health status, even when initial health is controlled for. However, the effects have opposite signs. Like previous investigators (e.g. Grossman, 1972a; Kemna, 1985), we find a very significant but **negative** impact of medical care use on health status. Clearly, this does not mean that health care has a large detrimental effect on health but rather that the visit frequency is an indicator of health deterioration. This finding casts serious doubt on the tenability of the derived demand hypothesis for the bulk of medical care use. Since our estimate cannot be a consequence of a negative bias due to the omission of initial health, it implies, again, a rejection of the basic Grossman model hypothesis that medical care use can be viewed as investment in good health.

Of course, the effects may look very different if very specific health measures and medical care interventions were considered but we are mainly interested in a fairly general, aggregate level model. We are not interested in testing whether indeed a course of penicillin has relieved the symptoms of a bacterial infection, but whether, on the

whole, people who do go to see a doctor more often are, *ceteris paribus*, better off after two years time. Perhaps the view of health care as investment is more appropriate for preventive rather than for curative services but we were unable to test this hypothesis⁴ due to lack of data on the reasons for physicians encounters.

Since a life-style judged to be healthy does seem to improve health status (or rather decrease deterioration), we now reach the remarkable conclusion that education has two **negative** indirect (allocative) effects on health status in 1983. Admittedly, these effects are very small and mostly not significant but they are in contradiction to what one would expect *a priori*.

Because the adequacy of our operationalizations of health input factors may be questioned, these findings are to be considered as tentative and requiring re-testing with better measures of the health investment inputs.

6.5. Conclusions

The general conclusion which emerges from the empirical analysis presented in this chapter is that schooling appears to have no marginal impact on health changes, i.e. on current health status when the health status of two years earlier is controlled for. The problem is that most of the other health determinants which show a clear 'cumulative' effect on health also have insignificant 'dynamic' effects. Notable exceptions are the influence of a healthy life-style and medical care utilization but these channel negative indirect effects of education on health. However, the decomposition analysis of education direct and indirect effects is hampered by the absence of any significant total (dynamic) effects.

When a statistical test does not indicate a difference, the conclusion is that either there is indeed no difference, or in other words, health profiles are parallel across levels of education; or the test has insufficient power to detect the differences. The latter interpretation is a real possibility here given the limitations of the dataset we analyzed. A sample size of 900 individuals between 21 and 65 years old is not large if changes in permanent health stock over a period of two years are the subject of analysis. The problem of limited dynamic variability is compounded by the latent variable approach: after correction for measurement error the correlation between the 'true health scores' in 1981 and 1983 is extremely high. As a consequence, the proportion of explained variance in 1983 latent health status is extremely high (84.1% in table 6.8) despite the fact that only a few variables have a statistically significant influence.

Fortunately, we have the opportunity to test the hypothesis that the

absence of a marginal health productivity effect of schooling may be due to a lack of discriminating power of the test that is being performed. This will be done in the next chapter.

Notes

1. Demonstration of the (over)identification of this and subsequent models is analogous to the procedure described in Appendix 4D. For a general exposition on the identification of LISREL models for the analysis of longitudinal data, see Jöreskog and Sörbom (1979), Chapter 5.
2. Because we were mainly interested in the equation explaining health in 1983, we used the 1983 values for the employment status variables. This may have caused (small) errors of classification of respondents for the 1981 equation. For example, it is possible that a small number of respondents were not yet retired, unemployed or disabled in 1981.
3. Some confidence in this subjective measure is obtained from the fact that the responses on HLIFEST correlate positively with physical exercising behaviour and negatively with smoking behaviour.
4. Edwards and Grossman (1983) found large significant impacts of a preventive dental visit on the oral health (periodontal index and the decay index) of 4121 adolescents but no effects at all of a physician checkup on measures of physical health.

Chapter 7. Use-related health depreciation and education: a fixed-effects model

"Even physicians and epidemiologists have a hard time specifying the production function for health. The commodity is produced probabilistically, substantially through the lifestyle choices of individuals." (Fuchs and Zeckhauser, 1987, p. 265)

7.1. Data

In this chapter, we analyse a different Dutch dataset which offers some advantages and some disadvantages compared to the one that was used in the previous chapter. These different characteristics both enable and require the use of a different method of analysis.

Since 1981, the Netherlands Central Bureau of Statistics (CBS) has been conducting a continuous Health Interview Survey. Each year a random sample of households is drawn from the general population excluding institutionalized persons. In principle, all residents at the selected addresses are interviewed, but since 1983 the maximum number of persons per household is four. The size of the response group amounts to about 10.000 persons a year. In addition to the continuous questionnaire administered each year, which includes information on health status, medical consumption and various personal background characteristics, the 1984 version contained some extra, (self-administered) retrospective questions about health and insurance status five years earlier. This feature made this essentially cross-sectional dataset also useful for a (limited) longitudinal analysis of individual health. The retrospective questions were only asked to the household members aged 16 years and over. This reduced the total size of the sample to about 7500 respondents.

The level of education was measured by two variables: highest level attended (EDUC1) and highest level obtained (EDUC2). In table 7.1 the

sample distributions for these two education variables are presented for comparison. It can be seen that a substantial part of the sample has followed a higher level of education than the one they also completed. This is only partly due to the fact that about 8% of the sample was still receiving education at the time of the interview.

Table 7.1. Sample frequency distribution according to three definitions of education level

	Code	Highest level completed	Highest level attended
Primary education	2	1854 (25.9)	911 (12.7)
Vocational training	3	2327 (32.5)	2706 (37.8)
Secondary education	4	2142 (29.9)	2402 (33.5)
Vocational College	5	632 (9.8)	908 (12.7)
University	6	206 (2.9)	234 (3.3)
TOTAL		7161 (100)	7161 (100)

The percentage distribution of sample respondents according to the level of education completed corresponds very well with the same distribution obtained by Van Ginneken and Appelboom (1986) for the combined CBS Health Interview Surveys of 1981 to 1984 ($n = 30351$). The sample can be said to be fairly representative for the educational distribution of the Dutch population aged 16 years and over. Before we explore the relationship between educational attainment and health in section 7.3, we will first discuss the way in which health (changes) were measured in this survey.

7.2. Measurement of health

The CBS Health Interview Survey contains similar questions on health as the ones that were used in the Nederland Oké Survey. However, for the retrospective health question a somewhat different type of scaling was used. Respondents were asked to rate their general health status as follows:

- (a) Can you give an appreciation mark for your state of health nowadays? (please circle a number)
- | | | | | | | | | | | |
|----------|---|---|---|---|---|---|---|---|----|-----------|
| very bad | | | | | | | | | | very good |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | |

(b) And how was your state of health about 5 years ago? (please circle a number)

very bad									very good
1	2	3	4	5	6	7	8	9	10

By asking these questions subsequently, people were not only invited to "mark" the level but also the change in their health status during the last 5 years. The mean response was 7.87 for the health mark in 1984 (HMARK84) and 8.02 in 1979 (HMARK79), which means that on average the health of this sample has deteriorated by 0.15 points on this (subjective) ten-point scale. The skewed distribution of responses around those means has slightly shifted to the left from 1979 to 1984. This can be seen from the row and column totals of the cross-tabulation presented in table 7.2. Although again the majority of the sample (65%) is located on the diagonal and therefore reports health to be stationary, some respondents also report important changes ranging from -9 to +9. It seems worthwhile to investigate whether this dynamic variability has some systematic component.

In Van den Berg and Van Doorslaer (1987) some simple reliability tests of the health appreciation marks were performed. It was shown that the average score for current health status (HMARK84) of five-year age groups did not significantly differ from the score for health status five years ago (HMARK79) of the five years older group. Therefore, there seemed to be no systematic bias due to "recall effects". It was also investigated whether respondents in the same category of self-rated health (SRH) gave different appreciation marks according to level of education. Also this hypothesis could be rejected: higher and lower education groups gave similar average appreciation figures for health within each SRH category.

Another important issue is the reliability and validity of this appreciation mark compared to other measures of good health. It resembles the self-rated health question (SRH), but it has more categories and has the additional connotation of a school mark in the Netherlands where 6 means 'sufficient' and 5 'insufficient'. In order to facilitate comparison with the latent health construct that was used in Chapter 6, we re-estimated a LISREL measurement model with HMARK84 as an additional fourth indicator, besides self-rated health on a 5-point scale (SRH), a dummy for the presence of at least one chronic condition or handicap (DCHRON) and the score on the 21-item complaints list (COMPL21). The estimation results are presented in table 7.3. Model I is the four-indicator analog of the measurement model used in Chapter 6.

It can be seen that the appreciation mark (HMARK84) performs very well: it has the highest reliability coefficient of all indicators. This is probably due to the fact that it shows somewhat more variability than the second best indicator, SRHB4. For example, whereas 79% of the respondents report their health status to be good or very good (codes 1 and 2 for SRHB4), about the same percentage of respondents is

Table 7.2. Cross-tabulation of health appreciation marks for 1984 and 1979 (CES, HIS, n = 7181)

		HMARK79										TOTAL	
		very poor											
		1	2	3	4	5	6	7	8	9	10		
HMARK84	very poor	1	8	0	3	1	1	2	1	3	0	4	22 (.3)
		2	3	4	0	5	3	4	1	1	1	0	22 (.3)
		3	2	1	11	5	8	5	5	8	4	7	56 (.8)
		4	1	2	4	31	9	21	22	21	9	9	128 (1.8)
		5	4	7	14	14	104	37	60	53	23	19	335 (4.7)
		6	4	2	12	28	34	182	135	134	42	34	607 (8.4)
		7	2	6	12	30	57	85	621	341	128	36	1317 (18.3)
		8	2	3	8	16	40	102	134	1575	229	90	2199 (30.6)
		9	0	4	3	7	12	23	44	88	1058	83	1322 (18.4)
	very good	10	2	2	3	4	6	8	12	44	35	1058	1173 (16.3)
TOTAL		28 (.4)	31 (.4)	70 (1.0)	140 (1.9)	274 (3.8)	468 (6.5)	1035 (14.4)	2268 (31.6)	1528 (21.3)	1340 (18.7)	7181 (100.0)	

Note: figures in parentheses are percentages

Table 7.3. Maximum likelihood estimation results of LISREL health measurement model with four indicators (n = 7150)

	I		II		III	
	$\hat{\lambda}$	$\hat{\rho}$	$\hat{\lambda}$	$\hat{\rho}$	$\hat{\lambda}$	$\hat{\rho}$
$y_1 = \text{HMARK84}$	1.00 (-)	.715	1.00 (-)	.728	1.00 (-)	.842
$y_2 = \text{SRHB4}$	-.569 (62.0)	.648	-.559 (59.7)	.634	-.486 (68.9)	.555
$y_3 = \text{COMPL21}$	-1.904 (55.9)	.467	-1.889 (55.2)	.468	-1.666 (58.3)	.421
$y_4 = \text{DCHRON}$	-.193 (28.3)	.129	-.182 (25.5)	.166	-.168 (27.9)	.115
$x_1 = \text{HMARK79 } (\hat{y}_{11})$	-		-		.606 (67.2)	
ψ_{11}	1.788 (39.0)		1.820 (38.8)		1.148 (39.9)	
θ_{e1}	.712 (28.1)		.680 (25.5)		.396 (19.4)	
θ_{e2}	.316 (35.0)		.327 (34.9)		.398 (48.0)	
θ_{e3}	7.385 (49.3)		7.374 (49.0)		8.026 (53.6)	
θ_{e4}	.449 (58.1)		.456 (57.5)		.457 (58.8)	
θ_{e42}			.027 (4.5)			
χ^2	21.0		.04		340	
df	2		1		5	
AGFI	.980		1.00		.796	

distributed across the four highest categories of HMARK84 (7 to 10). It can be said to be more sensitive to small variations in health states.

Despite the extremely high t-values for all indicators, model I is rejected by the χ^2 -test with two degrees of freedom. DCHRON has the lowest reliability.¹ The highest modification index in model I was found for θ_{e42} , the error covariance of the indicators SRHB4 and DCHRON. A very good fit ($\chi^2_{(1)} = .04$) was obtained when this parameter was not constrained to be zero in model version II. One possible reason for the statistical significance (in model II) of $\theta_{e42} = \text{cov}(e_4, e_2)$ is that these two indicators show much less variation than the two others. Because they are measured on the two most limited category scales (a two and a five-point scale), larger measurement errors are made by categorizing individuals according to their health.

Finally, we have related the latent, four-indicator, health index to

the value for the appreciation mark for health in 1979 (HMARK79) in order to obtain an estimate of the stability coefficient of health (γ_{11}) (see e.g. Wheaton, e.a., 1977). The estimated stability ($\hat{\gamma}_{11} = .606$) is much lower than in the analysis of the previous chapter.

We have tested whether this finding could be due to the fact that HMARK84 is corrected for measurement error while HMARK79 is not. As a consequence, the random fluctuations in HMARK79 could be larger than those in latent health η , which more or less equals HMARK84 corrected for measurement error. However, when the observed HMARK84 variable was regressed on HMARK79 we obtained $\hat{\gamma}_{11} = .627$ (t-value = 70.6) which is only slightly larger than γ_{11} in table 7.3. Therefore, we have to conclude that health status, as measured by HMARK at two points in time with five years in between is not as stable as the latent health status variable that was used in Chapter 6. In the next section we attempt to explain some of this variation.

7.3. Analysis of covariance

7.3.1. Dynamic effects

In Chapter 5 we have already indicated that the availability of longitudinal data allows for a more direct test of education effects on health than the estimation of demand for health functions. This direct approach essentially amounts to using the basic health change identity of the Grossman model for estimation. Recall that this identity can be written as:

$$H_t = H_t - H_{t-1} = I(t) - \delta(t, X) \cdot H_{t-1} \quad (7.1)$$

Education can either directly influence health investment $I(t)$ (Grossman's hypothesis) or have an influence through the depreciation rate δ (Maurinen's hypothesis). Theoretically, there is the distinction between a technical production and an allocative use-related effect but empirically it becomes a matter of distinguishing between first and second order effects in an equation like (7.1).

In our case, and if we take one period to be 5 years, H_t and H_{t-1} are directly observable but I_t is not. The CBS-HIS 1984 does not contain good measures of health investment which cover the same five year period. No information on healthy or unhealthy habits is collected and the (detailed) medical consumption during the last two months before the interview cannot reasonably be assumed to be a good proxy of medical care use over the entire five year period. Moreover, one would run into the same difficulties as in the previous chapter if one is unable to distinguish clearly between curative and preventive medical expenditure.

Therefore, we assumed health investment to be a direct function of a

number of personal background characteristics (income, education, age and sex), the practice adopted in those studies of health production functions where income is used as a proxy for health inputs (see Chapter 5 and 6). Of course, this simplification seriously weakens the test of the marginal productivity effect of education on health investment because it is now measured as the (reduced-form) effect on H_t and not as the (structural, but unobservable) effect on I_t . On the other hand, however, this approach has the advantage of enabling a direct estimation of δ and an exploration of its interaction with background characteristics. Comparison of the additive (or main) effects of these background variables with the interactive (or second-order) effects with initial health will provide more insight into the relative empirical importance of these two competing hypotheses.

We will approach the problem again in a stepwise manner by focusing first on the identification of main effects before introducing interaction. The size of the sample also permits testing for non-linearity of the background effects on health change. We, therefore, started off with an analysis of variance. Essentially, it does not make any difference whether H_t or \dot{H}_t is used as the dependent variable when H_{t-1} is among the explanatory variables.

Estimating

$$\dot{H}_t = H_t - H_{t-1} = \alpha + \beta H_{t-1} + \gamma X + \epsilon_t \quad (7.2)$$

is virtually equivalent to estimating

$$H_t = \alpha + (\beta') H_{t-1} + \gamma X + \epsilon_t \quad (7.3)$$

where $\beta' = \beta + 1$

and X = a vector of (time-invariant) exogenous variables

ϵ_t = a disturbance term

Of course, taking the first difference of the appreciation marks ($H = \text{HDIFF} = \text{HMARK84} - \text{HMARK79}$) implicitly assumes that these are measured on an interval scale, i.e. the scale of measurement is unique up to a linear transformation (see e.g. Van de Ven and Hooijmans, 1982). One could argue that an appreciation mark only has an ordinal or ranking scale, i.e. that the scale of measurement is only unique up to a monotonic transformation. This would mean that the difference between, say, a 3 and a 4 is not necessarily equal to the difference between a 9 and a 10. For the present analysis, we have made the simplifying assumption that health marks do represent a (categorical) interval scale. Future work should provide more insight into the scalability properties of health measures, perhaps also by experimenting with magnitude scaling techniques (see e.g. Lodge, 1981, Torrance, 1986).

Because we are mainly interested in finding out how much of the variance in HDIFF is attributable to education after the effects of other variables have been held constant, analysis of variance and analysis of covariance were considered to be appropriate techniques at the first stage. The questions we seek to answer are:

1. Which education variable is the most appropriate for our purposes?
2. Does education contribute to an explanation of the variance in HDIFF, and if so, what is the pattern of the effect?
3. How does the effect of education vary when other factors (= non-metric variables in SPSS terminology) or covariates (= metric variables in SPSS terminology) are controlled for?
4. Is there any interaction between the effect of education and other control variables, or between the control variables mutually?

In order to be able to answer these questions, a stepwise analysis of (co)variance was performed. Tables 7.4A to C summarize the results regarding the effects for each of the three education variables. Table 7.4A represents the effect of EDUC1 as the deviations from the overall sample mean for HDIFF (= -.15) for each of the five levels of education completed. The first row in each table contains the uncontrolled deviations from the mean. A remarkable finding is that we do not see a monotonically increasing schooling gradient but a curvilinear association: only the lowest and the highest education level show a health deterioration above the average. Especially the result obtained for the 206 academics (3% of the sample) is very unexpected. The pattern became much more pronounced when HMARK79 was controlled for in the next step: as a result, the gap between those with and without more than primary education widened, but the academics were still far below the level to be expected on the basis of the upward trend in all other levels of education. Nevertheless the standardized beta indicates a fairly strong positive effect of education. The F-test statistics indicate that not only the main (first-order) effect of education but also its interaction with HMARK79 is statistically significant ($p < .001$). This result already stands in sharp contrast to the previous chapter where no dynamic effect whatsoever was found for education when initial health was controlled for.

Because the individual's age is probably the most important 'confounder' in this relationship, it was introduced separately in the next step to obtain model A3. Indeed it was found that a large part of the effect of 'only primary education' could be attributed to age differences. Controlling for age did reduce but did not wipe out the significant effect of education on HDIFF. Because the effect of age as a categorical variable divided into 8 classes was found to be monotonically decreasing, the age variable was introduced as a (continuous) covariate in the next model in order to limit the number of different cells. Consequently, models A4 and B4 do not contain two-way interactions with age. Adding also sex and income (model A4) only slightly reduces the magnitude and significance of the education effect further. The increase in the significance of the interaction effect ($F(33)$ is now 2.6) may be due to the deletion of the interaction effects with age. We can conclude that EDUC1 has a significant but non-linear partial effect on HDIFF.

The results for the highest level of education attended (EDUC2) are

Table 7.4. Effects of education level on HDIFF expressed as deviations from the grand mean ($= -.15$) (ANOVA)

A. Highest level of education completed (EDUC1)**

Model	Control variables*	2						F(4) EDUC1	β EDUC1	F(33) EDUC1 #H79	R ² Total
		prim. educ.	3 voc. train.	4 sec. educ.	5 voc. coll.	6 univ.					
A1	uncontrolled	-.10	.02	.06	.06	-.13	-	.05	-	-	.0025
A2	health79	-.28	.06	.13	.16	.01	37.8	.13	3.2	3.2	.258
A3	health79, age	-.17	.03	.07	.14	.01	13.7	.08	1.9	1.9	.281
A4	health79, age, sex, income	-.14	.03	.06	.11	-.03	8.4	.06	2.6	2.6	.283
n		1854	2327	2142	632	206					

B. Highest level of education attended (EDUC2)

Model	Control variables	2						F(4) EDUC1	β EDUC1	F(33) EDUC1 *H79	R ² Total
		prim. educ.	3 voc. train.	4 sec. educ.	5 voc. coll.	6 univ.					
B1	no	-.21	-.00	.07	.07	-.10	-	.07	-	-	.005
B2	health79	-.49	-.03	.15	.18	.05	54.6	.15	4.8	4.8	.264
B3	health79, age	-.30	-.02	.08	.12	.04	17.6	.09	3.6	3.6	.282
B4	health79, age, sex, income	-.25	-.01	.07	.11	.00	11.5	.08	5.3	5.3	.283
n		911	2706	2402	908	234					

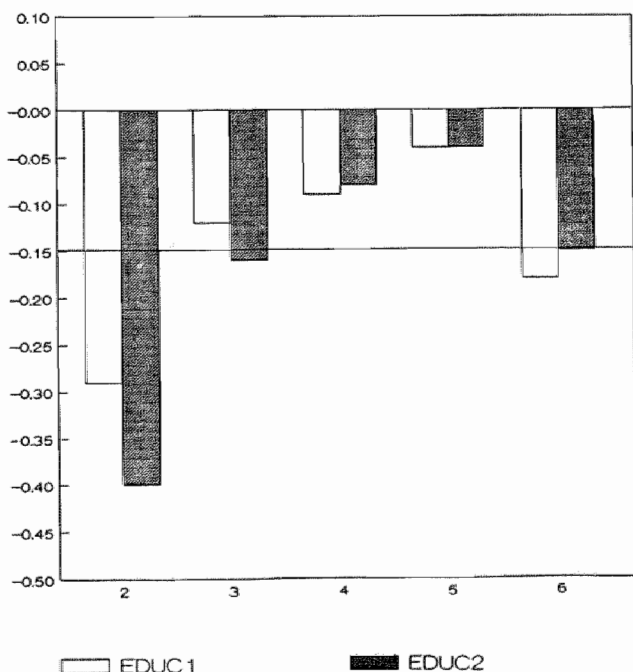
Notes:

*) The classical experimental approach is used in these model specifications, i.e. main effects are assessed while controlling for all other main effects but not interactions; in assessing interaction effects both all main effects and all other interaction effects are controlled for (Nie e.a., 1975, p. 414). In the fourth model, no interactions with age are included because it was entered as a covariate.

**) β is a standardized partial regression coefficient (Nie e.a., 1975, p. 417). Figures in parentheses with F statistics are degrees of freedom of numerator.

rather similar but with one important exception: by separating out those individuals who did not receive any more formal education beyond primary school, the differences at the lower end of the educational ladder become larger (see table 7.4B). This is the main reason for the higher statistical significance of the education effect. The position of the highest level of schooling (university) has slightly improved but is still substantially worse than that of levels 4 and 5. Figure 7.1 illustrates the differences between the partial effects of EDUC1 and EDUC2 graphically with two histograms. It is clear that by 'upgrading' those who did attend but not complete a higher level of schooling the lower education categories show a somewhat worse health status evolution.

Figure 7.1. Partial effect of EDUC1 (in model A4) and EDUC2 (in model B4) on HDIFF



Statistical significance of differences between separate levels of education is not tested by the analysis of (co)variance. This will be examined in the next section by means of a least-squares dummy variable approach to the fixed-effects model. However, the F-tests of the covariance analyses do bring out a clear statistical effect of all education variables and their interaction with initial health.

To illustrate the pattern and significance of the effects of the control variables, we have presented the full analysis of covariance

and multiple classification results of model B4 in table 7.5. From the decomposition of the total variation (= sum of squared (SS) deviations from the mean) in HDIFF into SS due to covariate, main and interaction effects, it can be seen that by far the largest amount of variation is explained by initial health (HMARK79).² There appears to be no difference in health change between males and females when other effects are controlled for. At the 0.05 level of statistical significance only the interactions between HMARK79 and EDUC2, and between HMARK79 and HHINC are significant.

Table 7.5A does not reveal the patterns of these effects on HDIFF. The latter are visualized in the multiple classification analysis presented in table 7.5B. It clearly illustrates how the small effect of SEX (female health depreciates .08 points more than male health) disappears after adjustment while most of the other main effects on the contrary become more pronounced after adjustment. HMARK79 has, of course, a strong negative impact on HDIFF and HHINC shows a small but consistent positive effect. Again, especially the lowest income

Table 7.5. Analysis of covariance and multiple classification results of model B4 (n = 7161)

A. ANALYSIS OF COVARIANCE

Source of variation	SS	DF	MS	F	P
AGE	56.84	1	56.84	44.5	< .001
Main effects	3676.75	20	183.84	144.0	< .001
HMARK79	3625.54	9	402.84	315.6	< .001
SEX	.21	1	.21	.2	.688
EDUC2	58.81	4	14.70	11.5	< .001
HHINC	38.24	6	6.37	5.0	< .001
Two-way interactions	413.19	131	3.29	2.6	< .001
HMARK79 x SEX	17.71	9	1.97	1.5	.129
HMARK79 x EDUC2	144.37	34	4.25	3.3	< .001
HMARK79 x HHINC	130.10	54	2.41	1.9	< .001
SEX x EDUC2	10.08	4	2.52	2.0	.096
SEX x HHINC	.47	6	.08	.1	.999
EDUC2 x HHINC	27.77	24	1.16	.9	.594
Explained	4164.79	152	27.40	21.47	< .001
Residual	8944.32	7008	1.27		
Total	13109.11	7160	1.83		

Note: SS = sum of squared deviations from the mean

DF = degrees of freedom

MS = SS divided by DF (mean squared)

F = MS divided by MS residual (= 1.27)

P = P-value associated with F statistic

B. MULTIPLE CLASSIFICATION ANALYSIS

(as deviations from sample mean HDIFF = -.15)

	n	unadjusted	eta	adjusted	beta
HMARK79			.49		.54
1 very poor	28	3.36		3.61	
2	31	4.21		4.42	
3	70	2.85		3.04	
4	173	2.07		2.25	
5	237	1.28		1.45	
6	465	.73		.87	
7	1034	.06		.12	
8	2263	-.18		-.19	
9	1526	-.34		-.41	
10 very good	1336	-.40		-.47	
SEX			.03		.00
1 male	3447	-.04		-.01	
2 female	3715	.04		.01	
EDUC2: see Table 7.4B, model B4					
HHINC (in guilders per year)			.04		.06
1 < 18,000	929	-.13		-.16	
2 18 - 22,000	934	-.03		-.05	
3 22 - 28,000	1222	.01		.01	
4 28 - 36,000	1096	.01		-.02	
5 36 - 45,000	946	.04		.09	
6 ≥ 45,000	824	.08		.09	
7 missing	1212	.02		.04	
AGE (model B3 results)			.09		.15
1 16 - 24 years	1356	.05		.22	
2 25 - 34	1590	.08		.17	
3 35 - 44	1364	.03		.05	
4 45 - 54	987	-.02		-.16	
5 55 - 64	864	-.02		-.18	
6 65 - 74	674	-.07		-.22	
7 75 - 84	280	-.39		-.49	
8 ≥ 85	47	-.93		-.84	
R² = .282					

category seems to be substantially worse off, even when age and education are controlled for. If income is seen as a proxy for health investment behaviour, this could be interpreted as a marginally beneficial influence of behaviour. Even if the effect were largely a result of environmental characteristics (housing conditions, recreation facilities, etc.), the significantly positive influence on health change over time is remarkable. However, it is dangerous to interpret the association entirely as an income effect on health

because household income cannot be assumed to be entirely exogenous.

Although AGE was entered as a covariate in model B4, we have also presented the results for the categorized age variable in model B3. Another remarkable finding is that health change is positive for the 16 to 24 years age category ($-.15 + .22 = .07$) and almost constant for the 25 to 34 age category ($-.15 + .17 = .02$) before health deterioration starts to increase beyond 35, reaching a maximum of $-.15 + (-.84) = -.99$ for the over 85 years old. This finding suggests an age profile of health stock which reaches its maximum around thirty years and depreciates at an increasing rate thereafter. It should be mentioned that the rate of health deterioration in the higher age groups is probably substantially underestimated due to sample selection bias (see Van den Berg and Van Doorslaer, 1987). The elderly whose health deteriorated most during 1979 and 1984, i.e. those who were admitted to hospital or even died are not represented in our sample.

Altogether, these five explanatory variables account for about 28% of the variation in HDIFF. Neither the proportion of total explained variance, nor the patterns of the effects of the individual variables differed much from the full model results for 4A (with EDUC1).

We can therefore conclude from this analysis of the dynamic variability of health by means of changes in appreciation marks that the lifecycle profiles of health are not parallel across education, age and income categories: health depreciates faster with rising age and falling education and income levels. Moreover, the effect of H_{t-1} (the depreciation rate δ) seems to depend on age and education, as suggested by the use-related depreciation hypothesis. Before we look into these interaction effects in more detail, we first analyze the cumulative (cross-sectional) effects in order to be able to compare with the results of the previous chapter.

7.3.2. Cumulative effects

Although we are not primarily interested in the cross-section results for the obvious reason that their causal interpretation is problematical, we do present them here for comparison with the previous chapter and with other explanatory studies of individual health differences. The dependent variable is now the absolute level of health in 1984 as measured by the appreciation mark HMARK84. The set of explanatory variables is the same as before but with H_{t-1} (= HMARK79) deleted from the equation. Table 7.6 presents the multiple classification results with the F-test statistics of the corresponding analysis of variance. The education variable used is the highest level attended (EDUC2) and age is introduced as a non-metric categorized factor.

Table 7.6. ANOVA and multiple classification results with HMARK84 as the dependent variable (n = 7165)

Grand mean = 7.87

Main effects	unadjusted deviations	eta	adjusted deviations	beta	F (df)	P
SEX		.05		.04	10.9	< .001
1 male	.09		.06			
2 female	- .08		- .06			
EDUC2		.22		.08	9.4 (4)	< .001
2 prim. educ.	- .82		- .31			
3 voc. train.	- .05		.00			
4 sec. educ.	.26		.07			
5 voc. coll.	.26		.08			
6 univ.	.23		.11			
HHINC		.14		.09	9.1 (6)	< .001
1 < 18,000	- .44		- .27			
2 18 - 22,000	- .24		.11			
3 22 - 28,000	.03		.02			
4 28 - 36,000	.05		- .05			
5 36 - 45,000	.26		.16			
6 ≥ 45,000	.22		.13			
7 missing	.09		.10			
AGE		.33		.29	79.5 (7)	< .001
1 16 - 24	.62		.58			
2 25 - 34	.44		.39			
3 35 - 44	.09		.03			
4 45 - 54	- .44		- .46			
5 55 - 64	- .51		- .46			
6 65 - 74	- .67		- .49			
7 75 - 84	- .98		- .76			
8 ≥ 84	-1.23		-1.00			
Two-way interactions				df	F	P
SEX x EDUC2				4	1.96	.097
SEX x HHINC				6	.15	.989
SEX x AGE				7	5.09	< .001
EDUC2 x HHINC				24	.82	.712
EDUC2 x AGE				28	1.32	.122
HHINC x AGE				42	1.61	.009

$R^2 = .123$

The analysis brings out the familiar relationships between health and the various background characteristics with lower health status for women, elderly, low income and lower education persons. In contrast to the dynamic analysis most of the adjusted associations are lower than the uncontrolled ones, which implies that the 'risk factors' (e.g. age, education and income) are not independently distributed. Perhaps

the most striking finding is that the adjusted deviations for EDUC2 are not much higher than in the dynamic analysis of table 7.4B except for the highest level. This means that the cumulative impact of education is not much higher than the dynamic effect, if at all. The deviations are now monotonically increasing with the level of education up until university level but the largest discrepancy remains between primary education and the others. The observed pattern of increasing health status with increasing level of education is consistent with the findings of Van Ginneken and Appelboom (1986) who described differences on other health indicators according to the highest level of schooling completed.

The health difference by SEX now remains significant also after adjustment. The .12 points lower appreciation mark for women is consistent with recently reported results on sex differentials in health and medical consumption based on the combined CBS HIS data of the years 1981 to 1984. Van den Brekel (1987) has shown that despite their higher life expectancy, Dutch women raise more health complaints, suffer more often from some chronic conditions and are more often underweight or obese than men. Only the interaction effects between age and sex and between age and income are significant at the .05 level of the F-test.

The main conclusion from the analyses of (co)variance reported in this section is that not only the level of education (attended) shows a significant dynamic effect on health status but that moreover it is almost as large as the cumulative effect, and it interacts significantly with previous health status. The former finding implies that, controlling for the health status of five years earlier is not as important for the effect of schooling on adult health as might be expected if both health and schooling were largely determined by a common cause (see Chapter 5). The largest part of the schooling-health association therefore does not seem to be spurious. The latter finding warrants a more detailed analysis of the pattern of interaction, especially with a view to test the use-related health depreciation hypothesis. This will be done in the next section.

7.4. Interaction effects

The ANCOVA results reported in the previous section showed that some two-way interaction effects were statistically significant according to an F-test. This meant that the effects of some variables were different for different groups. It could not be tested which groups exactly differed from which others and by how much. In this section, the objective is to unravel in more detail the interaction between age and education on the one hand and initial health status (five years ago) on the other. The fixed-effects model that will be specified is based on the information obtained from the ANCOVA results and will be

estimated using the least-squares dummy variable (LSDV) approach (see e.g. Hsiao, 1986). In principle, this approach is more general and it could also have been used in the previous section but it would have required the estimation of huge regression equations incorporating an enormous amount of dummy variables. These can now be simplified using the ANCOVA results. For instance, SEX will be left out of the analysis because of non-significance and AGE and HMARK79 will be introduced as continuous, metric variables because they both had consistent, negative effects.³ The income variable will be reduced to four 0/1 dummy variables.

First, we have re-estimated model B4, without the variable SEX or any interaction terms, by means of OLS. The estimation results are presented in the first column of table 7.7. The first conclusion to be drawn from these results is that the coefficient of HMARK79 (-.42) is very close to $\beta' - 1$, where $\beta' = .60$ is the estimated stability parameter in model III of table 7.3. Income, age and education show similar effects as in the analysis of covariance and the coefficients of all dummies are significantly different from the reference category.

Table 7.7. OLS estimation results of regression models for HDIFF with and without interaction terms (n = 7158)

Explanatory variables	Model I	Model II	Model III
Intercept	3.405 (31.8)*	2.974 (37.2)	2.387 (7.6)
HHINC 18 - 22000	.108 (1.97)	.114 (2.07)	.106 (1.93)
HHINC 22 - 36000	.138 (2.96)	.157 (3.35)	.140 (2.99)
HHINC \geq 36000	.213 (4.24)	.238 (4.75)	.216 (4.31)
HHINC missing	.189 (3.62)	.207 (3.96)	.191 (3.66)
HMARK79 (= H79)	-.421 (46.8)	-.366 (31.9)	-.284 (7.35)
AGE	-.013 (14.0)	-	-.0073 (1.60)
EDUC2 = 3	.203 (4.20)	-	1.264 (6.37)
EDUC2 = 4	.260 (4.99)	-	1.124 (5.0)
EDUC2 = 5	.275 (4.52)	-	1.465 (4.83)
EDUC2 = 6	.188 (2.06)	-	.133 (.24)
Interaction with H79:			
AGE	-	-.0015 (13.8)	-.0007 (1.26)
EDUC2 = 3	-	.017 (2.76)	-.141 (5.53)
EDUC2 = 4	-	.025 (3.75)	-.116 (4.10)
EDUC2 = 5	-	.026 (3.43)	-.156 (4.17)
EDUC2 = 6	-	.019 (1.72)	-.006 (.10)
R ²	.239	.237	.242
F	225.6	221.6	153.5
df**	10	10	15

*) Absolute t-values in parentheses

**) df of numerator, (df of denominator of F-test is always larger than 7100)

Next we estimated a model II in which age and education are assumed to have only an indirect effect through the coefficient of HMARK79 (δ) on HDIFF. The pattern which emerges is as expected: a negative influence of age and a positive influence of education. Because the effect of EDUC2 again is not strictly increasing with each level but decreases at the university level, we also performed a test for the statistical significance of the differences between subsequent levels. By redefining the dummy variables for education as minimal levels (greater than or equal) we obtained the results presented in table 7.8. It can be seen that the largest (and most significant) difference is between the two lowest levels of education 2 and 3 (primary education and vocational training). After that, there is only a small increase from 3 to 4 (secondary level) but the latter does not differ anymore from the higher levels 5 and 6. The decrease from 5 to 6 is not statistically significant. On the whole, the evidence indicates decreasing health returns to additional education.

Table 7.8. OLS estimation results for models I and IV with increasing dummy variables for education

	MODEL I	MODEL II (interaction)
EDUC2 \geq 3	.203 (4.20)*	.0172 (2.76)
EDUC2 \geq 4	.057 (1.69)	.0077 (1.87)
EDUC2 \geq 5	.015 (.32)	.0012 (.22)
EDUC2 \geq 6	-.088 (1.01)	-.0072 (.70)

*) absolute t-values in parentheses

Finally, we considered the possibility that there may be both main effects on HDIFF and interaction effects via HMARK79 for age and education by estimating model III. This caused some spectacular changes in the results. Both the main and interaction effects of the AGE and EDUC2 variable have changed quite dramatically. Age no longer has a significant direct nor interaction effect on HDIFF. Because the simple correlation between AGE and AGE * HMARK79 equals .85, collinearity may, to some extent, contribute to an explanation of the large estimated standard errors for the coefficients of these variables. Nevertheless, the total effect of AGE, evaluated at the mean of HMARK ($-.0073 - .0007 * 7.87 = .0128$), is very close to the effects that were estimated in models I and II.

In contrast to the AGE effect, the total positive effect of the EDUC dummies is now not decomposed into two smaller positive effects, but into a larger positive main effect and a negative interaction effect with initial health (H79). Probably as a result of multicollinearity (simple correlations between the EDUC2 dummies and their interaction

variable with H79 vary between .96 and .98) their estimated standard errors have increased. But because this increase is smaller than the increase (in absolute value) of the estimated coefficients, the t-values for three of the four dummies are higher than in the separate models I and II. The only exceptions are again the dummy variable for university level (EDUC2 = 6) and its interaction with H79. Both have now very small coefficients and very large standard errors.

Furthermore, because models I and II are nested in model III, we could test the validity of restricting either the main effects or the interaction effects of AGE and EDUC2 to be zero by means of an F test for coefficient restrictions. The following F test was used to test the null hypothesis that some of the coefficients in model III are equal to zero, i.e. that either the main or interaction dummies for AGE and EDUC2 could be deleted from the equation (Kmenta, 1971, p. 371)

$$\left| \frac{R^2_Q - R^2_K}{1 - R^2_Q} \right| \left| \frac{n - Q}{Q - K} \right| \sim F_{Q-K, n-Q}$$

where R^2_Q = the coefficient of determination of the full model (with Q variables)

R^2_K = the coefficient of determination of the restricted model (with K variables)

n = the number of observations

Q-K = the number of restricted coefficients

Since for both the restricted models $Q-K = 5$ and $n-Q = 7142$, the tabulated critical value of $F_{5,\infty}$ at the 1% level of significance is 3.02. Both the F statistic for the model I restrictions ($F = 7.6$) and for the model II restrictions ($F = 13.2$) exceed this critical value and therefore these hypotheses can be rejected. Apparently, the overall positive effect of education on HDIFF in models I and II conceal two opposite effects of which the positive dominates the negative one.

In a graphical representation of the (negative) relationship between health change (HDIFF) and initial health (H79) the three models have very different implications for the regression line differences between high and low education groups. These are illustrated in Figure 7.2. Subdividing in two levels of education - high (HE) and low (LE) - and abstracting from the peculiar results for the highest level of EDUC2, the results can be described in terms of intercept and slope differences, i.e.:

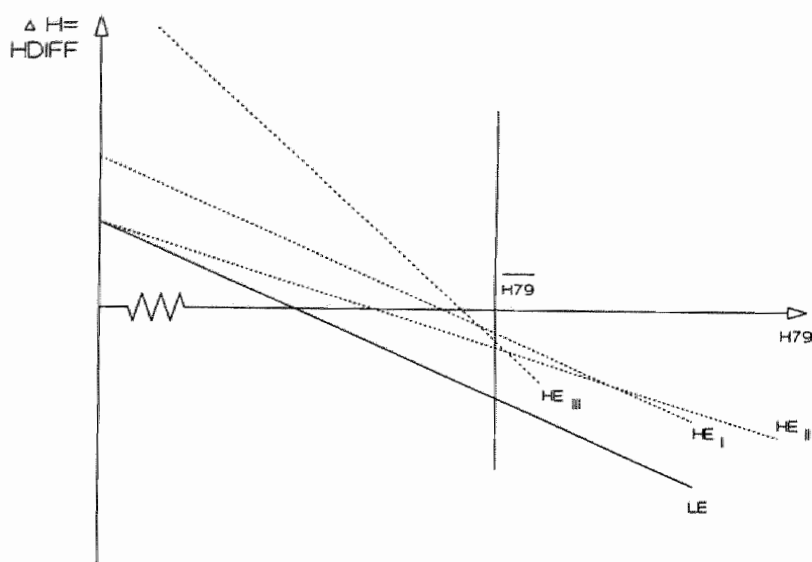
Model I: parallel lines but HE higher intercept

Model II: same intercept but LE steeper slope and lines diverging with rising initial health

Model III: HE higher intercept but also steeper slope and regression lines converging with rising initial health.

The graph clearly illustrates the biases introduced by restricting either the intercepts or the slopes of the two educational categories to be equal. On the other hand, it has to be noted that around the

Figure 7.2. Simplified graphical representation of different schooling effects in model I, II and III of table 7.7



mean initial health status ($H79$), where the majority of observations is located, the differences in health change by educational group are very similar. Interpreted literally in terms of the health change equation (7.1), model III implies that the (positive) productivity-increasing effect dominates a (negative) use-related depreciation effect. However, due to the weak operationalization of health investment this inference is very premature. A replication of this test with more adequate measures of health investment than household income is required before one can be more conclusive about this interpretation.

7.5. Conclusions

In this chapter we have examined empirically the dynamic relationship between education and health by means of a dataset which differs substantially from the one we used in Chapter 6. Because also the results are quite different from the earlier ones, it is worthwhile to summarize the most important differences between the two samples. This is done in table 7.9.

We think that most of these differences are crucial to understand why in Chapter 6 we did not, and in Chapter 7 we did find a significant

Table 7.9. A summary of characteristics of the Nederland Oké Panel and the CBS H.I.S.

	Nederland Oké	CBS H.I.S.
Sample size	895	7158
Age group	21 - 65	16+
Reference period (health)	1981 - 1983	1979 - 1984
Time dimension	panel	retrospective questions
Health indicators used	latent variable based on three indicators (SRH, CHRON and COMFL)	appreciation mark (1 to 10)
Educational levels distinguished	4	5

dynamic impact of education on health change over time. The main reason is undoubtedly an enormous gain in test power in the second analysis. This is not only due to the fact that the sample size is eight times larger but also to the larger variation in both the dependent (health change) and the independent (education) variable. Clearly, extending the reference period from two to five years and including the people over 65, substantially increases the dynamic variation in health, but also the retrospective health rating on a ten-point scale seems to be a more sensitive measurement instrument. Moreover, most of the health effect of education seems to be located at the bottom of the educational ladder. In that respect, the educational subdivision used in Chapter 6 is rather unfortunate because it lumps together fifty percent of the sample into the lowest education category.

For the reasons mentioned above, we believe that we were unable to identify a significant dynamic health effect of education in Chapter 6 mainly because of a lack of test power. The analysis of the 1984 CBS Health Interview Survey reported in this chapter shows that:

- a. Lifecycle profiles of health are not parallel across groups of people with different levels of education. Especially the health status of those who did not attend more than primary education depreciates faster than that of the other classes. A result, which was not strong enough to be conclusive but consistent enough to be cause concern, was the negative marginal health returns to university education. No intervening factors could be examined in detail which could contribute to an explanation of this finding. One interesting finding was that this education effect did not differ between those who did and those who did not have a paid job.⁴ This makes an explanation based on working conditions unlikely.
- b. On average, health deterioration only starts beyond the age of 30 but increases at an increasing rate thereafter. The acceleration

occurs particularly after 65 and this is in line with the theoretical expectations of Muurinen: "If the impact of age on the rate of depreciation is seen as an exponentially increasing function of age, both the finiteness of life in a biological sense is guaranteed and also all these other (use-related depreciation) effects will ultimately become negligible in relation to the age-induced decrease in health." (Muurinen, 1982a, p. 7)

- c. By lack of a good measure of health investment during the five year period considered, household income was used as a proxy, which co-determines investment behaviour. A strongly positive effect on health change was found, irrespective of the specification used to estimate the relationship. It means that in the Netherlands not only the absolute health level is positively related to income (cfr. e.g. Van Vliet and Van Praag, 1987; Van Ginneken and Appelboom, 1987) but also the rate of change when initial health is controlled for.
- d. In contrast to the previous findings the effect of gender on absolute health differences was **not** replicated by the dynamic analysis. When especially initial health is held constant, no sex differences in health change were found: male health profiles are slightly higher than - but parallel with - female health profiles.
- e. Appreciation marks are fairly reliable measurement instruments for general health status. They are sufficiently sensitive to analyze longitudinal differences.
- f. Neither Grossman's marginal productivity-hypothesis nor Muurinen's use-related depreciation hypothesis about the effect of education could be rejected. When estimated separately, both the direct main effect and the second-order interaction effect with previous health were statistically significant and had the expected sign. However, when both types of effects were introduced simultaneously the interaction effect changed sign which indicates that the total or overall effect conceals two opposite effects. The latter finding is considered to be an interesting topic for further analysis with better measures of health investment.

Notes

1. When the dummy DCHRON was replaced by the number of chronic conditions or handicaps, the reliability coefficient increased to .336. This is consistent with the fairly high reliability coefficient obtained for this indicator by Van Vliet and Van Praag (1987).
2. A general exposition on the interpretation of the ANCOVA-tables obtained with the SPSS program can be found in e.g. Tabachnick and Fidell (1983), Chapter 7.
3. We do not consider non-linear relationships with these variables here.
4. The ANCOVA results referred to were not presented but were obtained by replacing the SEX dummy by a JOB (0/1) dummy in model B4. The interaction between JOB and EDUC2 was not significant while the other effects hardly changed.

Chapter 8. General summary and conclusions

"Great theories are rarely simple in the ordinary sense of the term. Both quantum mechanisms and relativity theory are very difficult to understand; it takes only a few minutes to memorize the facts accounted for by relativity, but years of study may not suffice to master the theory and to see these facts in its context." (Polanyi, 1958, p. 16)

In the introduction to this thesis, we stated the central research question as follows: How, and to what extent, does the knowledge capital of individuals affect their behaviour with respect to health and with respect to their demand for medical care? Because the explicit intention was to approach this problem from an economic point of view, we first reviewed the theoretical models that had been used in the health economics literature in the (distant and recent) past in Chapter 2. The demand models that were discussed were classified according to whether or not:

1. the demand for medical care was derived from the more fundamental demand for 'good health',
2. health was regarded as a human capital stock which depreciates over time but can be augmented by health investment,
3. health depreciation was considered as uncertain, thereby causing a demand for insurance and the need to distinguish between preventive and curative care, and
4. consumer information on the effectiveness of medical care was imperfect, leading to a derived demand for (professional) advice.

Although undoubtedly each of these model characteristics is important, and therefore ought to be incorporated in any attempt to construct a comprehensive theoretical model of health (care) demand behaviour, such an exercise was not our ambition. Because we mainly wanted to use the theoretical models as the basis for our empirical analysis, we chose to tackle the problem in two stages for pragmatic reasons. A static but stochastic model was chosen to study the relationship between consumer information or knowledge and the demand for medical

care in Part I. The empirical analysis in the second part is based on a dynamic but deterministic theoretical model of the demand for health. Abstracting from intertemporal considerations in the first part, and from uncertainty in the second part, reduced the research problem to manageable proportions without, in our view, running the risk of oversimplification.

Part I: Consumer information and the demand for medical care

In Chapter 3 basically Kihlstrom's (1974a) Bayesian approach to the demand for information when product quality is uncertain, is adopted to describe demand behaviour in the health care market. The essence of the idea is that consumers are uncertain about the quality of medical care in terms of what it can contribute to their health status. It is assumed that they do possess some prior knowledge which is characterized by an a priori (normal) distribution of their belief about the effectiveness of medical care. The larger the variance around the mean of this distribution (or the more diffuse the prior), the more uncertain they are and the more valuable the availability of additional information. Information can be obtained from experience but it can also be bought in the 'medical advice' market. Basically, diagnostic and therapeutic information is precisely what is being purchased in most (primary care) physician-patient encounters.

It is further assumed that patients act as Bayesian decision makers in combining their prior knowledge and the advice received (the sample information) to arrive at a posterior knowledge (in the form of an a posteriori distribution) on which to base their decisions in order to maximize expected utility. In such a framework, the degree of consumer ignorance (and thus sovereignty) varies continuously with the degree of uncertainty expressed by the prior and posterior distributions. It is then shown that the expected utility gains of extra information (obtained from a doctor-expert or elsewhere) and treatment will be weighted against the cost of information. Some comparative static results can be derived for the effects of income, prices, prior uncertainty and the perceived informativeness of medical advice on the demand for diagnostic and therapeutic information.

Additional problems arise in health care because a doctor visit very often is a joint product consisting of both information and care. The Bayesian decision making assumption provides a key to the issue of how suppliers of medical care may be able to induce additional demand for their services: by altering the accuracy of their advice and overstating the marginal effect of an intervention. The consumer's prior knowledge and his perception of the accuracy of such advice provide limits to this discretionary power of the physician. How effective these limits are is an empirical matter.

This brings us to Chapter 4 in which a (LISREL) structural equation model is specified in order to test some of the theoretical model predictions. The dependent variables are two patient-initiated demands for primary medical care (self-medication and GP contacts) and two physician-initiated utilization rates (referred GP contacts and prescription drugs). A sample of nearly a thousand adults was analyzed drawn from the general Flemish population.

Special attention was paid to the measurement of health and knowledge. Individual health status was measured as a latent vector consisting of two components: a permanent stock of health and a transitory flow of healthy time. In a similar way, two Bayesian components of consumer information were distinguished: a MIMIC index for medical knowledge-ability was constructed to measure the prior information, and by means of principal components analysis a scale was derived to measure the respondents' perception of the accuracy of GP advice. The reliability of the various indicators we used and the validity of the constructs we developed was analyzed extensively in the measurement submodels. The results of these health and information 'modules' proved to be very robust when they were integrated into the full structural model. This indicates that it was justified to adopt the stepwise strategy of sorting out the measurement problems before performing any hypothesis testing in the total model.

According to the estimation results of the basic model specification reported in table 7.13, the following conclusions can be drawn with respect to the tested model predictions. First of all, individuals with better medical information do not initiate fewer visits to their GP. On the contrary, *ceteris paribus*, they report significantly more of such contacts but the effect is quite small. Its total (positive) effect is even smaller than the direct effect, mainly because knowledge has also a positive influence on the use of non-prescribed medicines. Because this acts as a substitute for GP contacts, the indirect effect is negative.

Distinguishing between the initiator of consumption turned out to be important because the opposite effect was found for physician-initiated GP contacts. People with more medical knowledge had significantly less GP contacts on referral. It was argued that this finding, in combination with the consistently positive, though non-significant effect of GP availability, could be interpreted as evidence in favour of Pauly's (1980) hypothesis that prior information effectively constrains the providers' ability to induce demand for their services. It does not, however, rule out other possible explanations.

No influence was found of prior medical knowledge on the use of prescription drugs and not one of the demands we considered was significantly affected by our measure of the (in)accuracy perception of GP diagnostic and therapeutic information. Education only showed a positive influence on the use of self-medication.

In the discussion of these results in Chapter 4 we pointed out that their inconsistency with some of the major theoretical predictions stands in sharp contrast to their consistency with earlier findings in the literature. Either no or a small positive effect was found of medical knowledge on physician visits. However, with our structural model approach it was possible to show that some of the interpretations which other researchers had given to this result were incorrect. In addition, we could show that the opposite result holds for physician-initiated contacts.

Our general conclusion of Part One is therefore that the Bayesian approach to the demand for information, despite its theoretical appeal, contributes little to a better empirical explanation of the demand for primary care. Even if the power of the tests could be raised by analyzing a larger sample or measuring more accurately than we did, some effects may become statistically more significant but it is unlikely that they will be much larger. The institutional requirement to buy professional information, in the form of a prescription, for the majority of health care may be one of the reasons why the effect of prior information is so small. In fact, such requirements may always cause a downward bias in the information effect, even when in future research efforts it is attempted to distinguish between more and less appropriate medical consumption.

Finally, also the estimated partial direct effects of medical knowledge and education on health status were very small and often non-significant. This finding is inconsistent with the strong partial correlations between the level of schooling and health which are often found also in individual cross-section studies. This relationship is explored in more detail in the second part of this study.

Part II: Education and health

The second part of the study does not (only) deal with the specific mechanism through which knowledge might affect medical consumption but takes a broader view to investigate the overall effect of one type of human capital (education) on the other type of human capital (health). The theoretical and empirical problems associated with an analysis of the schooling-health relationship are discussed in Chapter 5. It was argued that, for our purposes, education effects on the health of a general adult population could better be examined by means of production functions rather than demand functions. By estimating an essentially technical relationship like a production function, a number of specific problems inherent to the demand for health approach can be avoided. At the same time, however, most of the characteristics of the economic approach to health behaviour are neglected.

Through a review of earlier empirical studies of the schooling-health

link, we identified mainly four methodological problems: (a) omitted variable bias, (b) measurement error, (c) direct versus indirect effects, and (d) non-linearity and interaction. With respect to the issue of omitted variable bias, it was shown that longitudinal data might provide an acceptable solution if the objective is not so much to identify the "omitted third factor(s)" but rather to control for possible spurious correlation. The problem of errors in health measurement can be reduced by using several indicators of the unobservable variable and combining them into one index. This procedure has hitherto only been used in (non-longitudinal) cross-section studies. Investigations of intervening factors in the linkage between schooling and health sometimes reported important indirect effects through allocation decisions. But the problem with these estimates is that possibly the total effect is substantially overestimated due to lack of control for omitted variables. Finally, few studies allowed for the possibility of non-linear and/or non-additive effects of education.

For our own analysis of education effects in health production, we had two datasets available, none of which allowed us to tackle all of these methodological problems simultaneously. However, each of the two had strengths and weaknesses compared to the other. Therefore, we could not discard any of them a priori and analyzed both separately.

In Chapter 6 a longitudinal latent variable model is estimated with data taken from the two waves of the Nederland Oké Panel Survey. The health status of the 895 individuals in the sample in 1981 and 1983 is measured as a latent variable with three observable indicators: self-rated general health status, the presence of a chronic condition or handicap and a 21-item psycho-somatic complaints score. By allowing for intertemporal correlation between the error terms of two of these indicators, a measurement model was obtained that fitted these data. A recursive structural equation model is then specified to estimate the cumulative effects of education and other exogenous variables on 1981 health status and the dynamic effects on 1983 health status. The main conclusion is that schooling appears to have no marginal impact on health changes, i.e. on current health when the health status of two years earlier is controlled for. Both education and most of the other health determinants show a clear "cumulative" effect but insignificant "dynamic" effects. Notable exceptions are the influence of the two health inputs, i.e. a healthy lifestyle and the number of doctor visits, but these channel negative indirect effects of education on health. However, a decomposition into direct and indirect effects of education is hampered by the absence of any significant total (dynamic) effects. This would imply that the health lifecycle profiles of the more educated have a higher intercept but identical slopes, i.e. they are parallel. However, we have some doubt about the generalizability of this finding because the results also indicate that age does not have a dynamic effect on health, i.e. that health depreciation does not rise with age. Even though the sample only

includes persons between the age limits of 21 and 65, one would expect some age impact on health change.

We therefore conclude that the absence of a marginal health productivity effect of schooling (and other variables) may be due to a lack of discriminating power of the test. By far the largest part of the high percentage (84%) of explained variance in 1983 latent health status is accounted for by the health status two years earlier. Both the limited variation of the latent health variable over a period of two years and the limited size of the sample may prevent the detection of significant dynamic effects. This hypothesis could be tested in Chapter 7 with the other dataset.

More recently another Dutch dataset has become available with which the dynamic relationship between health and education could be examined. The 1984 CBS Health Interview Survey is not a panel but it includes a retrospective health question. Respondents were asked to rate both their current and past health status (five years ago) by means of an appreciation mark between 1 and 10. This was found to be a fairly reliable and sufficiently sensitive measure to analyze longitudinal changes in health. The longer reference period (five years), the larger sample size ($n = 7158$) and the better subdivision of education categories increased the power of the test substantially.

By means of an analysis of covariance of the health changes over time, we found that lifecycle profiles of health are not parallel across groups of people with different levels of schooling. In particular, the health depreciation rate of those who did not attend more than primary education is higher than that of the other classes (except university level), even when income, age, sex and initial health are controlled for. Although no intervening factors could be examined in detail, an explanation based on working conditions seems unlikely because this education effect did not differ between those who did and those who did not have a paid job. Further investigation is needed of the specific mechanisms through which this effect materializes. People with a low education level do seem to continue to be a high health risk group deserving the attention of health policy makers. Our analysis indicates that differences in education levels increase health inequalities, thereby amplifying the initial inequality.

Not only the direct main effect of education on health change but also the interaction effect with initial health status was statistically significant. Both indicated a substantial health gain between the lowest and the second level of education but little health returns to additional education beyond vocational training. When both types of effects were introduced simultaneously, the interaction effect became negative. This result suggests that the positive overall effect of education would conceal two opposite effects rather than two complementary influences, as one would expect. It certainly contradicts the use-related depreciation hypothesis about the effect

of education which could not be rejected when the interaction effects were estimated separately (without the main effects). The marginal productivity hypothesis could not be tested adequately due to the lack of a good measure of health investment. This will be a necessity in future work if one wants to discriminate more adequately between the two competing hypotheses.

With respect to the effects of the other background characteristics, a non-linear age influence was found: health deterioration only starts beyond the age of 30 but increases at an increasing rate thereafter. Male health profiles are slightly higher but parallel with female health profiles: the initial sex difference remains constant over the lifecycle. In all estimates, household income showed a strongly positive partial effect on health which again indicates that initial socio-economic inequalities grow as people grow older.

The main general conclusion which emerges from the various analyses presented in the second part of this book is that the correlation between health and education is not spurious. More schooling does seem to have a beneficial causal effect on health or, formulated negatively, a lack of education adversely affects the expected time path of individual health status. Future research will have to concentrate on more specific causal mechanisms which may channel this influence. In order to identify how much of this effect can be attributed to working conditions, lifestyle, use of medical care or other intervening variables, preferably panel data should be used but according to our analysis, also retrospective questions in cross-section surveys provide a useful and less expensive alternative source of longitudinal information.

References

- Acton, J.P. (1985), Nonmonetary factors in the demand for medical services: some empirical evidence, *Journal of Political Economy* 83, pp 595-614
- Adriaanse, H., M.J. Drop, R. Halfens and H. Philipsen (1981), *Leeft Nederland Oke ? Verslag van een onderzoek naar de beleving, de opvattingen en de gedragingen inzake gezondheid*, Stichting Nederland Oke, Zeist
- Andersen, R., J. Kravits and O.W. Anderson (1975) (eds), *Equity in health services*, Ballinger Publ. Co., Cambridge, Massachusetts
- Arrow, K.J. (1963), Uncertainty and the welfare economics of medical care, *American Economic Review*, 53, pp 941-973, also reprinted in M. Cooper and A.J. Culyer (eds), *Health economics*, Penguin Books, Harmondsworth, 1973
- Auster, R.D., I. Leveson and D. Sarachek (1969), The production of health: an exploratory study, *Journal of Human Resources* 4, pp 412-436
- Becker, G.S. (1965), A theory of the allocation of time, *Economic Journal* 75, pp 493-517
- Bentler, P.M. and D.G. Bonett (1980), Significance tests and goodness of fit in the analysis of covariance structures, *Psychological Bulletin* 88, No. 3, pp 588-606
- Blaug, M. (1970), *An introduction to the economics of education*, Penguin Books, New York
- Blaug, M. (1980), *The methodology of economics, or: how economists explain*, Cambridge University Press, Cambridge
- Boomsma, A. (1983), *On the robustness of LISREL (maximum likelihood estimation) against small sample size and non-normality*, Ph.D. dissertation, Groningen, Sociometric Research Foundation, Amsterdam
- Brekel, E.J.G. van den (1987), *Gezondheid en medische consumptie naar geslacht*, Maandbericht Gezondheidsstatistiek CBS, 6, 2, pp 5-55
- Breyer, F. (1982), Rational purchase of medical care and differential insurance coverage for diagnostic services, *Journal of Health Economics* 1, pp 147-156
- Brook, R.H., J.E. Ware, A. Davies-Avery et al. (1979), Overview of adult health status measures fielded in Rand's Health Insurance Study, *Medical Care* 17 (Suppl.), pp 1-131
- Carrin, G. and J. Van Dael (1984), An empirical model of the demand for health care in Belgium, *Applied Economics*, Vol. 3, pp 317-334
- Chamberlain, G. (1977), Education, income and ability revisited, in: D.J. Aigner and A.S. Goldberger (eds), *Latent variables in socio-economic models*, North-Holland, Amsterdam
- Coyte, P.C. (1983), The economics of Medicare: equilibrium within the medical community, *Journal of Labor Economics* 1, 3, pp 264-285
- Coyte, P.C. (1985), The market for medical services and physicians: an application of hedonic price theory, *Canadian Journal of Economics* 18, 2, pp 377-394
- Cropper, M.L. (1977), Health investment in health and occupational choice, *Journal of Political Economy* 85, 6, pp 1273-1294
- Cropper, M.L. (1981), Measuring the benefits from reduced morbidity, *American Economic Review* 71, 2, pp 235-240
- Ollis, J. and P. West (1979), *The economics of health: an introduction*, Martin Robertson, Oxford
- Culyer, A.J. (1971), Is medical care different? In: M. Cooper and A.J. Culyer (eds), *Health Economics*, Penguin Books, Harmondsworth

DeGroot, M.H. (1970), *Optimal statistical decisions*, McGraw-Hill, New York

Dejardin, J. and F. Praet (1980), Invloed van de verhoogde tegemoetkoming van de verplichte ziekteverzekering op de kosten en consumptie, *Belgisch Tijdschrift voor Sociale Zekerheid*, nr. 2 (Ministerie van Sociale Voorzorg, Brussel)

Dirken, J.M. (1967), *Arbeid en stress*, proefschrift, Wolters, Groningen

Doorslaer, E.K.A. Van (1984), Effects of cost-sharing on the demand for prescription drugs in Belgium, *Acta Hospitalia*, No. 3, pp 69-81

Dorfman, R. (1969), An economic interpretation of optimal control theory, *American Economic Review*, 59, pp 817-831

Edwards, L.N. and M. Grossman (1983), Adolescent health, family background and preventive medical care, in: D. Salkever, I. Sirageldin and A. Sorkin (eds) *Research in human capital and development*, Vol. 3, JAI Press, pp 77-109

Eeckhoudt, L., T. Lebrun and J.C. Salliy (1985), Risk aversion and physicians' medical decision-making, *Journal of Health Economics* 4, No. 2, pp 273-281

Ehrlich, I. and G.S. Becker (1972), Market insurance, self-insurance and self-protection, *Journal of Political Economy*, 80, 4, pp 623-648

Evans, R.G. (1976), Book review of: M. Perlman (ed.), (1974), *The economics of health and medical care*, London, MacMillan, in: *Canadian Journal of Economics* 9, 3, pp 532-537

Evans, R. (1984), *Strained mercy. The economics of Canadian health care*, Butterworths, Toronto

Farrell, P. and V.R. Fuchs (1982), Schooling and health: the cigarette connection, *Journal of Health Economics* 1, pp 217-230

Feldstein, M.S. (1974), *Econometric Studies of Health Economics*. In: M.D. Intriligator and D.A. Kendrick (eds), *Frontiers of quantitative economics*, Vol. 2, North-Holland, Amsterdam

Foets, M., F. Berghmans and L. Janssens (1975), The primary health care project in Belgium: a survey on the utilization of services. *Social Science and Medicine*, Vol. 20, No. 3, pp 181-190

Freixas, X. and R.E. Kihlstrom (1984), Risk aversion and information demand, in: M. Boyer and R.E. Kihlstrom (eds) *Bayesian models in economic theory*, Elsevier Science Publications, pp 93-104

Fuchs, V.R. (1974), *Who shall live? Health, economics and social choice*, Basic Books, New York

Fuchs, V.R. (1979), The economics of health in a postindustrial society, *The Public Interest*, pp 3-20

Fuchs, V. (ed.) (1982), *Economic aspects of health*, University of Chicago Press, NBER Conference Report, Chicago

Fuchs, V.R. (1986), *The Health Economy*, Harvard University Press, Cambridge, Massachusetts

Gaag, J. van der and B. Wolfe (1983), *Estimating the demand for medical care: health as a critical factor for adults and children*. Applied Econometrics Association, Lyon

Ginneken, J.K. van and W.J. Appelboom (1986), Influence of socio-economic factors on perceived morbidity and the use of health services in the Netherlands, *Maandbericht Gezondheidsstatistiek CBS* 5, 10, pp 5-9

Goldberger, A.S. (1974), Unobservable variables in econometrics, in: P. Zarembka (ed.) *Frontiers of econometrics*, Academic Press, New York, pp 193-213

- Goldman, F. and M. Grossman (1978), The demand for pediatric care: an hedonic approach, *Journal of Political Economy* 86, 2, pp 259-280
- Griliches, Z. (1974), Errors in variables and other unobservables, *Econometrica*, 42, also reprinted in: D.J. Aigner and A.S. Goldberger (eds) *Latent variables in socio-economic models*, North-Holland, Amsterdam
- Griliches, Z. and W.M. Mason (1972), Education, income and ability, *Journal of Political Economy* 80, pp 74-103
- Grossman, M. (1972a), *The demand for health: a theoretical and empirical investigation*, National Bureau of Economic Research, New York
- Grossman, M. (1972b), On the concept of health capital and the demand for health, *Journal of Political Economy* 80, 2, pp 223-255
- Grossman, M. and E.H. Rand (1974), Consumer incentives for health services in chronic illness, in: S.J. Mushkin (ed.) *Consumer incentives for health care*, Prodist, New York, pp 114-151
- Grossman, M. (1975), The correlation between health and schooling. In: N.E. Terleckyj (ed.) *Household production and consumption*, Columbia University Press, New York, pp 147-211
- Grossman, M. (1982), The demand for health after a decade, *Journal of Health Economics*, 1, 1, pp 1-3
- Grossman, S., R. Kihlstrom and L.J. Mirman (1977), A Bayesian approach to the production of information and learning by doing. *Review of Economic Studies*, 44, pp 533-547
- Halfens, R., M.J. Drop and H. Philipsen (1984), *Leefwijzen en subjectieve gezondheid van een panel uit de Nederlandse bevolking*, Rijksuniversiteit Limburg, Maastricht
- Haveman, R.H. and B.L. Wolfe (1984), Schooling and economic well-being: the role of non-market effects, *Journal of Human Resources*, Vol. 19, No. 3, pp 377-407
- Hay, J. and M.J. Leahy (1982), Physician-induced demand. An empirical analysis of the consumer information gap. *Journal of Health Economics* 1, pp 231-244
- Heckman, J. (1976), A lifecycle model of earnings, learning and consumption, *Journal of Political Economy* 84, pp 11-44
- Hey, J.D. and M.S. Patel (1983), Prevention and cure or: is an ounce of prevention worth a pound of cure? *Journal of Health Economics* 2, pp 119-138
- Hey, J.D. (1979), *Uncertainty in micro-economics*, Martin Robertson, Oxford
- Hirshleifer, J. and J.G. Riley (1979), The analytics of uncertainty and information. An expository survey. *Journal of Economic Literature* 17, pp 1375-1421
- Hooymans, E. and W. van de Ven (1982), Implementing a health status index in a structural health care model. In: J. van der Gaag et al. (eds) *Economics of health care*, Praeger Grants Economics Series, pp 302-322
- Hsiao, C. (1986), *Analysis of panel data*, Cambridge University Press, New York
- Ippolito, P.M. (1981), Information and the life cycle consumption of hazardous goods, *Economic Inquiry* 19, pp 529-558
- Johnston, J. (1972), *Econometric methods*, McGraw-Hill, New York
- Jöreskog, K. (1973), A general method for estimating a structural equation system, in: Goldberger, A.S. and D. Duncan (eds) *Structural equation models in the social sciences*, Seminar Press, New York, pp 1-18

- Jöreskog, K.G. and A.S. Goldberger (1975), Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association*, Vol. 70, pp 631-639
- Jöreskog, K. and D. Sörbom (1977), Statistical models and methods for analysis of longitudinal data, in: D.J. Aigner and A.S. Goldberger (eds) *Latent variables in socio-economic models*, North-Holland, Amsterdam, pp 285-325
- Jöreskog, K.G. and D. Sörbom (1979), *Advances in factor analysis and structural equation models*, Abt Books, Cambridge, Massachusetts
- Jöreskog, K.G. and D. Sörbom (1981), LISREL version V. *Analysis of linear structural relationships by the method of maximum likelihood*. User's guide international Educational Services, Chicago
- Kamien, M. and N. Schwartz (1981), *Dynamic optimisation: the calculus of variations and optimal control in economics and management*, North-Holland, Amsterdam
- Keeler, E.B., J.P. Newhouse and C.E. Phelps (1977), Deductibles and the demand for medical care services: the theory of a consumer facing a variable price schedule under uncertainty, *Econometrica* 45, 3, pp 641-655
- Kenna, H.J.M.I. (1987), Working conditions and the relationship between schooling and health, *Journal of Health Economics*, Vol.6, No. 3, pp 165-281
- Kenna, H.J.M.I. (1985), *The relationship between schooling and health: investigation into occupational linkage*, Ph.D. Dissertation, University of Pittsburgh
- Kihlstrom, R. (1974b), A general theory of demand for information about product quality. *Journal of Economic Theory* 8, pp 413-439
- Kihlstrom, R. (1974a), A Bayesian model of demand for information about product quality. *International Economic Review*, Vol. 15, No. 1, pp 99-118
- Kmenta, J. (1971), *Elements of econometrics*, MacMillan, New York
- Kohn, R. and K.L. White (1976), *Health care. An international study*, Oxford University Press, London
- Lancaster, K.J. (1966), A new approach to consumer theory, *Journal of Political Economy*, 74, pp 132-157
- Lee, L.F. (1982), Health and wage: a simultaneous equation model with multiple discrete indicators. *International Economic Review*, Vol. 23, pp 199-221
- Leigh, J.P. (1983), Direct and indirect effects of education on health, *Social Science and Medicine* 17, 4, pp 227-234
- Leu, R.E. and R. Doppman (1984), The demand for health care in Switzerland - a latent variable approach. In: W. Van Elteren et al. (eds) *Third International Conference on Systems Science in Health Care*, Springer-Verlag, Berlin
- Levine, D.S. and D. Yett (1973), A method for constructing proxy measures of health status. In: R.L. Berg (ed.) *Health Status Indexes*, Hospital Research and Educational Trust, Chicago
- Lodge, M. (1981), *Magnitude scaling: quantitative measurement of opinions*, Sage Publications, Series on quantitative applications in the social sciences, No. 25, Beverly Hills and London
- Mackenbach, J.P. and P.J. van der Maas (1987), Sociale ongelijkheid en verschillen in gezondheid. Een overzicht van de belangrijkste onderzoeksbevindingen, in: Wetenschappelijke Raad voor het Regeringsbeleid, *De ongelijke verdeling van gezondheid*, Serie Voorstudies en Achtergronden, nr. V58, Staatsuitgeverij, Den Haag
- MacCurdy, T.E. (1981), An empirical model of labor supply in a lifecycle setting, *Journal of Political Economy* 89, pp 1058-1085

- Maddala, G.S. (1979), *Econometrics*, McGraw-Hill, New York
- Manning, W.G. Jr., J.P. Newhouse and J.E. Ware, The status of health in demand estimation; or, beyond Excellent, Good, Fair, Poor. In: V. Fuchs (ed.) *Economic aspects of health*, NBER, University of Chicago Press, pp 143-184
- Markus, G.B. (1979), *Analyzing panel data*. Sage Publications, Series on quantitative applications in the social sciences, No. 18, Beverly Hills and London
- Michael, R.T. (1972), Education in nonmarket production, *Journal of Political Economy* 80, pp 306-327
- Mushkin, S. (1962), Health as investment, *Journal of Political Economy* 70, pp 129-157
- Murthen, B. (1983), Latent structural equation modeling with categorical data, in: J. de Leeuw, W.J. Keller and T. Wansbeek (eds), (1983), *Interfaces between econometrics and psychometrics*, *Annals of Applied Econometrics*, a supplement to *The Journal of Econometrics*, Vol. 22, pp 43-65
- Murinen, J.M. (1982), Demand for health: a generalised Grossman model. *Journal of Health Economics*, Vol. 1, No. 1, pp 5-28
- Murinen, J.M. (1982b), *An economic model of health behaviour with empirical applications to Finnish health survey data*, Ph.D. Dissertation, University of York, York
- Nelson, P. (1970), Information and consumer behaviour, *Journal of Political Economy* 78
- Newhouse, J.P. (1978), *The economics of medical care*, Addison Wesley
- Newhouse, J.P. (1979), The erosion of the medical market place, in: R.M. Scheffler(ed), *Research in health economics*, Vol. 1, JAI Press, Connecticut
- Newhouse, J.P. (1981), The demand for medical care services: a retrospect and prospect, in: J. van der Gaag and M. Perlman (eds) *Health, economics and health economics*, North-Holland, Amsterdam
- Newhouse, J.P. and L.J. Friedlander (1980), The relationship between medical resources and measures of health: some additional evidence, *Journal of Human Resources* 15, No. 2, pp 200-218
- Newhouse, J.P., C.E. Phelps and M.S. Marquis (1980), On having your cake and eating it too: Econometric problems in estimating the demand for health services, *Journal of Econometrics* 13, pp 365-390
- Newhouse, J.P., J.E. Ware and C. Donald (1981), How sophisticated are consumers about the medical care delivery system? *Medical Care*, Vol. 14, No. 3, pp 316-328
- NOTA 2000. Over de ontwikkeling van gezondheidsbeleid: feiten, beschouwingen en beleidsvoornemens, Tweede Kamer, vergaderjaar 1985-1986, 19500, nrs 1 en 2
- Nuyens, Y. (1980), *De eerste lijn is krom*, Van Loghum Slaterus, Deventer
- Organisation for Economic Cooperation and Development (1985), *Measuring health care 1960-1983. Expenditure, costs and performance*, Social Policy Studies, No. 2
- Pauly, M.V. and M. Satterthwaite (1981), The pricing of primary care physician's services: a test of the role of consumer information. *The Bell Journal of Economics*, Vol. 12, pp 488-506
- Pauly, M.V. (1980), *Doctors and their workshops: economic models of physician behavior*, NBER, University of Chicago Press
- Pauly, M.V. (1978), Is medical care different? In: W. Greenberg (ed.) *Competition in the health care sector*, Aspen Systems Corp., Germantown

- Phelps, C.E. (1973), **The demand for health insurance: a theoretical and empirical investigation**, R-1054-OED, The Rand Corporation, Santa Monica
- Phelps, C.E. (1976), Demand for reimbursement insurance, in: R.N. Rosett (ed.) **The role of health insurance in the health services sector**, National Bureau of Economic Research, New York
- Phelps, C.E. (1978), Illness prevention and medical insurance, **Journal of Human Resources** 16, Supplement, pp 183-207
- Phelps, C.E. and J.P. Newhouse (1974), Coinsurance, the price of time and the demand for medical services, **Review of Economics and Statistics** 56, pp 334-342
- Polanyi, M. (1958), **Personal Knowledge. Towards a post-critical philosophy**, Routledge & Kegan Paul, London
- Robinson, P.M. and M.C. Ferrara (1977), The estimation of a model for an unobservable variable with endogenous causes. In: D.J. Aigner and A.J. Goldberger (eds) **Latent variables in socio-economic models**, North-Holland, Amsterdam
- Rossiter, L.F. and G.R. Wilensky (1983), A reexamination of the use of physician services: the role of physician-initiated demand, **Inquiry**, Vol. 20, pp 162-172
- Saris, W.E., J. den Ronden and A. Satorra (1985), Testing structural equation models, in: R.F. Cuttance and J.R. Ecos (eds) **Structural modelling**, Cambridge University Press, New York
- Saris, W.E., A. Satorra and D. Sörbom (1986), **The detection and correction of specification errors in structural equation models**, Sociometric Research Foundation, Amsterdam
- Saris, W. and H. Stronkhorst (1984), **Causal modelling in non-experimental research**. Sociometric Research Foundation, Amsterdam
- Shakotko, R.A., L.N. Edwards and M. Grossman (1981), An exploration of the dynamic relationship between health and cognitive development in adolescence, in: J. van der Gaag and M. Perlman (eds) **Health economics and health economics**, North-Holland, Amsterdam
- Sindelar, J.L. (1982), Behaviorally caused loss of health and the use of medical care, **Economic Inquiry** 20, pp 458-471
- Tabachnick, B.G. and L.S. Fidell (1983), **Using multivariate statistics**, Harper and Row, New York
- Taubman, P. and S. Rosen (1982), Healthiness, education and marital status, in: V.R. Fuchs (ed.) **Economic aspects of health**, The University of Chicago Press, Chicago
- Telser, L. (1981), The market for research and development: physician demand and drug company supply, in: R. Helms (ed.), **Drugs and health**, American Enterprise Institute, Washington D.C.
- Torrance, G.W. (1986), Measurement of health state utilities for economic appraisal: A review, **Journal of Health Economics**, Vol. 5, No. 1, pp 1-30
- Usher, D. (1975), Comments on "The correlation between health and schooling". In: N.E. Terleckyj (eds) **Household production and consumption**, Columbia University Press, New York, pp 212-220
- Ven, W. van de and J. van der Gaag (1982), Health as an unobservable. A MIMIC-model of the demand for health care, **Journal of Health Economics** 1, 2, pp 157-183
- Ven, W.P.M.M. van de (1987), The key role of health insurance in a cost-effective health care system. Towards regulated competition in the Dutch health care market. **Health Policy** 7, pp 253-272

- Ven, W.P.M.M. van de and B.M.S. van Praag (1981), The demand for deductibles in private health insurance. A Probit model with sample selection, *Journal of Econometrics*, 17, pp 229-252
- Ven, W. van de and E. Hooymans (1982), The MIMIC health status index: what it is and how to use it. Report 82.17, Center for Research in Public Economics, Leyden University
- Viscusi, W. Kip and C.J. O'Connor (1984), Adaptive responses to chemical labeling: are workers Bayesian decision makers? *American Economic Review* 74, 5, pp 942-956
- Vliet, R.C.J.A. van and B.M.S. van Praag (1987), Health status estimation on the basis of MIMIC-health care models, *Journal of Health Economics* 6, No. 1, pp 27-42
- Wagstaff, A. (1986a), The demand for health: some new empirical evidence, *Journal of Health Economics* 5, pp 195-233
- Wagstaff, A. (1986b), The demand for health: a simplified Grossman model, *Bulletin of Economic Research* 38, 1, pp 93-95
- Welch, F. (1970), Education in production, *Journal of Political Economy* 78, pp 35-59
- Wennberg, J.E. et al. (1982), Professional uncertainty and the problem of supplier-induced demand. *Social Science and Medicine*, Vol. 16, pp 811-824
- Wheaton, G., B. Muthen, D.E. Alwin and G.F. Summers (1977), Assessing reliability and stability in panel models with multiple indicators, in: D.R. Heise (ed.) *Sociological Methodology*, 1977, San Francisco, Jossey-Bass, pp 84-136
- Wiley, D.E. (1973), The identification problem for structural equation models with unmeasured variables, in: A.S. Goldberger and O.D. Duncan (eds) *Structural equation models in the social sciences*, New York, Seminar Press, pp 69-83
- Wolfe, B. and J. van der Gaag (1981), A new health status index for children. In: J. van der Gaag and M. Perlman (eds) *Health economics and health economics*, North-Holland, Amsterdam
- Wolfe, B. and J. Behrman (1984), Determinants of women's health status and health care utilization in a developing country: a latent variable approach, *The Review of Economics and Statistics*, 56, Nov., pp 696-703

Appendices

APPENDIX 4A

Description of variables from the Primary Health Care Survey (n = 949)

VAR.NAME	DESCRIPTION	MEAN	VARIANCE	SKEWNESS
Health care utilization variables				
SMED	number of different types of non-prescribed self-medication used during 9 weeks (max. = 8)	.803	1.232	1.80
GPOON	number of contacts with GP during 9 weeks at patient's initiative (max. = 12)	.575	1.242	3.53
GPREF	number of contacts with GP during 9 weeks at physician's initiative (max. = 13)	.162	.535	9.13
PMED	number of different types of prescribed medicines used during 9 weeks (max. = 11)	1.105	2.911	2.03
Health status variables				
RAD(L)	number of restricted activity days during 9 weeks (max. = 63)	.309	.685	2.84
NAD(L)	number of non-activity days or in bed during 9 weeks (max. = 63)	.312	.727	2.87
OCOND	number of chronic conditions or handicaps of respondent at the time of the interview (max. = 6)	0.669	1.063	1.90
COMPL	number of different health complaints expressed by the respondent at the time during 9 weeks (max. = 3)	1.641	2.112	1.04
ACUTE	'acute' (flu-cold, infectious disease, trauma, digestive, NTE complaint categories)	0.459	0.519	1.74
CHRON	'chronic' (heart, pain, psychosocial illness, headache, rheumatic, sleeping problems, other)	0.667	1.022	2.29
DISAB(L)	'disability index' based on the frequency (0 = never, 1 = sometimes, 2 = often, 3 = always) with which the respondent (during 9 weeks) - had been unable to perform daily work - had been unable to practice normal hobbies or recreation - had felt himself more worried or less happy than normal - had not slept very well (in. = 0, max. = 26) the index is the sum of the score at the 3 interviews	0.877	0.940	.68
Knowledge variables (answer categories 0 = wrong, 1 = do not know, 2 = correct)				
TEST1	sum of scores on 14 items of general theoretical medical knowledge	16.453	16.889	.32
TEST2	sum of scores on 16 items of theoretical medical knowledge with some practical relevance (9 to 32)	21.121	15.638	.08
TEST3	sum of scores on 11 items of practical medical knowledge (2 to 22)	11.664	14.261	.19
TEST4	sum of scores on 2 items of knowledge of health care organization (0 to 4)	2.896	1.166	-.45

VAR.NAME	DESCRIPTION	MEAN	VARIANCE	SKEWNESS
Background variables				
SEX	1 = female	0.445	0.247	.22
AGE	min. = 16, max. = 90 years	42.198	286.862	.45
EDUC	education, 4 categories (0 to 3)	0.997	1.117	.61
SES	socio-economic status based on occupation of respondent (or partner, or head of household), 13 categories	3.854	11.733	.43
HHSZ	size of the household (1 to 11)	3.416	2.309	.74
FINC	monthly net-income of the household (0 to 3), 4 categories	1.713	1.025	-.27
PSPR	number of psycho-social problems of respondent at the time of the interview (max. = 3)	0.195	0.223	2.67
IHIST(L)	illness history of respondent, measured by weighted sum of chronic conditions or handicaps the respondent had suffered from in the past (max. = 32)	.762	.940	.82
MEDFAM	number of household members with medical or para-medical profession	0.075	0.228	8.11
WOPI	a dummy for widows, invalids, pensioners and orphans which comprise a category of sickfund insured with lower deductibles	0.156	0.132	1.90
GPDENS	GP density at municipal level (per 10,000 population)	5.692		
PHARD	pharmacy density at municipal level (per 10,000 population)	4.296	4.070	.17
Principal component factors				
INACPT		-.284	1.991	.01
INACP2	see text	3.788	2.023	.00
ACCURP2		4.516	1.859	-.31

(L) This means that the variable is transformed according to $\ln(\text{VAR} + 1)$ in order to reduce skewness.

APPENDIX 4B

Correlation matrix used as input for the model presented in table 4.13

	1	2	3	4	5	6	7	8	9	10
	SMED	GPOON	GPREF	FMED	COOND	PGPC	COMPL	LRAD	LNAD	LDISAB
1	1.000									
2	0.110	1.000								
3	0.017	0.057	1.000							
4	0.115	0.567	0.323	1.000						
5	0.054	0.239	0.172	0.426	1.000					
6	0.061	0.389	0.207	0.485	0.348	1.000				
7	0.605	0.350	0.153	0.519	0.292	0.332	1.000			
8	0.183	0.266	0.137	0.306	0.252	0.209	0.284	1.000		
9	0.085	0.389	0.180	0.407	0.204	0.228	0.259	0.228	1.000	
10	0.182	0.368	0.212	0.446	0.296	0.323	0.413	0.434	0.489	1.000
11	0.129	-0.009	-0.082	-0.042	-0.085	-0.049	0.067	-0.022	-0.031	-0.001
12	0.105	0.027	-0.040	-0.023	-0.040	-0.040	0.075	0.003	-0.045	0.013
13	0.114	-0.007	-0.099	-0.031	-0.088	-0.043	0.046	-0.001	-0.052	-0.003
14	-0.001	-0.009	-0.043	-0.045	-0.012	-0.049	0.055	0.020	0.035	0.056
15	-0.005	-0.041	-0.017	-0.081	-0.080	-0.115	-0.032	-0.048	-0.006	-0.036
16	0.138	-0.109	-0.099	-0.146	-0.173	-0.122	0.028	-0.108	-0.064	-0.064
17	0.145	-0.089	-0.059	-0.089	-0.127	-0.095	0.048	-0.038	-0.063	-0.016
18	0.054	-0.033	-0.005	-0.037	-0.018	-0.056	0.033	0.009	-0.023	0.026
19	0.010	-0.080	-0.062	-0.193	-0.231	-0.146	-0.100	-0.051	-0.062	-0.119
20	-0.106	0.133	0.120	0.295	0.365	0.209	0.075	0.141	-0.007	0.107
21	0.178	0.065	0.051	0.119	0.079	0.139	0.169	0.088	-0.068	0.107
22	0.069	0.099	0.024	0.094	0.135	0.098	0.136	0.081	0.153	0.234
23	0.036	0.045	0.046	0.140	0.067	0.157	0.146	0.081	0.028	0.097
24	0.074	-0.121	-0.101	-0.169	-0.260	-0.155	-0.036	-0.118	-0.036	-0.031
25	-0.034	0.166	0.099	0.254	0.310	0.193	0.098	0.150	0.062	0.081
26	0.045	0.028	0.068	0.058	0.087	0.081	0.108	0.003	0.053	0.072
27	0.021	0.002	0.041	-0.013	-0.064	-0.053	0.032	0.007	0.014	0.032
	11	12	13	14	15	16	17	18	19	20
	TEST1	TEST2	TEST3	TEST4	INACTP	EDUC	SES	MEDFAM	HHSZ	AGE
11	1.000									
12	0.506	1.000								
13	0.502	0.462	1.000							
14	0.227	0.252	0.200	1.000						
15	0.105	0.106	0.105	0.059	1.000					
16	0.437	0.351	0.370	0.187	0.120	1.000				
17	0.416	0.340	0.317	0.143	0.088	0.633	1.000			
18	0.154	0.162	0.203	0.017	0.088	0.224	0.162	1.000		
19	0.076	0.070	0.129	0.038	0.031	0.147	0.124	0.021	1.000	
20	-0.207	-0.150	-0.234	-0.127	-0.164	-0.404	-0.175	-0.130	-0.317	1.000
21	0.021	0.129	0.043	-0.006	-0.032	-0.035	0.003	0.073	0.020	-0.018
22	-0.014	-0.023	-0.006	0.029	0.015	-0.026	-0.033	-0.037	-0.018	0.015
23	0.002	0.019	-0.027	0.015	-0.009	-0.050	-0.032	0.014	-0.117	0.140
24	0.292	0.291	0.224	0.135	0.093	0.411	0.456	0.090	0.290	-0.364
25	-0.158	-0.145	-0.185	-0.107	-0.061	-0.235	-0.189	-0.061	-0.265	0.510
26	0.061	0.081	0.018	0.036	-0.002	0.074	0.121	0.046	-0.138	0.046
27	0.000	-0.026	-0.025	-0.024	-0.037	0.003	0.030	0.017	0.010	0.032
	21	22	23	24	25	26	27			
	FEM	PSPR	LIHIST	FINC	WOPI	PHARD	GPDENS			
21	1.000									
22	-0.001	1.000								
23	-0.096	0.094	1.000							
24	-0.011	-0.026	-0.103	1.000						
25	0.060	0.019	0.087	-0.426	1.000					
26	0.007	0.040	0.116	0.056	0.011	1.000				
27	-0.044	0.022	-0.005	-0.013	-0.045	0.158	1.000			

Survey items about evaluation of general practitioner services (diagnostic accuracy and therapeutic effectiveness)

NAME	STATEMENT	FREQUENCY DISTRIBUTION OF ANSWERS (n = 1655)			
		AGREE	PART. AGREE	NO OPINION	PART. DISAGREE
EVGP 1	- GP's do not use enough technical equipment (apparatus) to find out what is wrong	366	177	426	330
EVGP 2	- GP's listen patiently when you talk to them about small inconveniences	149	225	228	428
EVGP 3	- GP's spend enough time to examine you thoroughly when you are ill	189	233	179	461
EVGP 4	- When you go to the GP, you are always being helped	174	321	170	533
EVGP 5	- GP's make too little use of laboratory tests to find out what the problem is	351	200	425	321
EVGP 6	- GP's make sure that you are not being submitted to unnecessary tests	145	224	481	416
EVGP 7	- You must have a very special health problem if GP's do not know what is wrong	179	223	360	368
EVGP 8	- No matter how long a GP has studied, in many cases he does not know what is wrong	242	225	363	432
EVGP 9	- GP's give their patients sufficient explanation about their illness	336	306	236	423
EVGP10	- GP's do not examine you thoroughly enough	386	241	272	444
EVGP11	- GP's give you too little opportunity to ask questions	462	263	233	372
EVGP12	- In many cases, GP's hope for good luck when trying to cure you	490	224	532	247

APPENDIX 4D

Identification of the basic structural equation model presented in tables 4.12 and 4.13
(by E.S. Tan)

1. The model can be written as follows:

A measurement model:

$$\begin{aligned}
 \cdot Y_1 &= \eta_1 \\
 \cdot Y_2 &= \eta_2 \\
 \cdot Y_3 &= \eta_3 \\
 \cdot Y_4 &= \eta_4 \\
 \cdot Y_5 &= -\eta_5 + \epsilon_5 \\
 \cdot Y_6 &= \lambda_1 \eta_5 + \epsilon_6 \\
 \cdot Y_7 &= \lambda_2 \eta_5 + \epsilon_7 \\
 \cdot Y_8 &= \lambda_3 \eta_6 + \epsilon_8 \\
 \cdot Y_9 &= \lambda_4 \eta_6 + \epsilon_9 \\
 \cdot Y_{10} &= -\eta_6 + \epsilon_{10} \\
 \cdot Y_{11} &= \eta_7 + \epsilon_{11} \\
 \cdot Y_{12} &= \lambda_5 \eta_7 + \epsilon_{12} \\
 \cdot Y_{13} &= \lambda_6 \eta_7 + \epsilon_{13} \\
 \cdot Y_{14} &= \lambda_7 \eta_7 + \epsilon_{14} \\
 \cdot Y_{15} &= \eta_8
 \end{aligned}$$

and a structural model:

$$\begin{aligned}
 \cdot \eta_1 &= \beta_{16} \eta_6 + \beta_{17} \eta_7 + \beta_{18} \eta_8 + \gamma_{11} x_1 + \gamma_{13} x_3 + \gamma_{15} x_5 + \gamma_{16} x_6 + \gamma_{19} x_9 + \gamma_{1,10} x_{10} + \xi_1 \\
 \cdot \eta_2 &= \beta_{21} \eta_1 + \beta_{26} \eta_6 + \beta_{27} \eta_7 + \beta_{28} \eta_8 + \gamma_{21} x_1 + \gamma_{23} x_3 + \gamma_{25} x_5 + \gamma_{26} x_6 + \gamma_{29} x_9 + \gamma_{2,10} x_{10} \\
 &\quad + \xi_2 \\
 \cdot \eta_3 &= \beta_{32} \eta_2 + \beta_{35} \eta_5 + \beta_{37} \eta_7 + \beta_{38} \eta_8 + \gamma_{31} x_1 + \gamma_{33} x_3 + \gamma_{35} x_5 + \gamma_{36} x_6 + \gamma_{39} x_9 + \gamma_{3,10} x_{10} \\
 &\quad + \xi_3 \\
 \cdot \eta_4 &= \beta_{41} \eta_1 + \beta_{42} \eta_2 + \beta_{43} \eta_3 + \beta_{45} \eta_5 + \gamma_{47} x_7 + \gamma_{48} x_8 + \gamma_{41} x_1 + \gamma_{43} x_3 + \gamma_{45} x_5 + \gamma_{46} x_6 \\
 &\quad + \gamma_{49} x_9 + \gamma_{4,10} x_{10} + \xi_4 \\
 \cdot \eta_5 &= \beta_{57} \eta_7 + \gamma_{51} x_1 + \gamma_{53} x_3 + \gamma_{54} x_4 + \gamma_{55} x_5 + \gamma_{56} x_6 + \gamma_{57} x_7 + \gamma_{58} x_8 + \xi_5 \\
 \cdot \eta_6 &= \beta_{65} \eta_5 + \beta_{67} \eta_7 + \gamma_{61} x_1 + \gamma_{63} x_3 + \gamma_{64} x_4 + \gamma_{65} x_5 + \gamma_{66} x_6 + \gamma_{67} x_7 + \gamma_{68} x_8 + \xi_6 \\
 \cdot \eta_7 &= \beta_{75} \eta_5 + \gamma_{71} x_1 + \gamma_{72} x_2 + \gamma_{73} x_3 + \gamma_{74} x_4 + \gamma_{75} x_5 + \gamma_{76} x_6 + \xi_7 \\
 \cdot \eta_8 &= \beta_{87} \eta_7 + \gamma_{81} x_1 + \gamma_{82} x_2 + \gamma_{83} x_3 + \gamma_{84} x_4 + \gamma_{85} x_5 + \gamma_{86} x_6 + \xi_8
 \end{aligned}$$

Ψ = the variance-covariance matrix of the ξ 's is diagonal.

All covariances between the ϵ 's equal zero except $\text{cov}(\epsilon_5, \epsilon_6) \neq 0$, $\text{cov}(\epsilon_8, \epsilon_9) \neq 0$

We can split the whole model into three parts (submodels):

1. a measurement model,
2. a simultaneous structural model, and
3. a recursive structural model

2. The first part specifies the following measurement model:

$$Y = \Lambda \eta + \epsilon, \text{ with}$$

$$Y = (Y_1, Y_5, Y_6, Y_7, Y_8, Y_9, Y_{10}, Y_{11}, Y_{12}, Y_{13}, Y_{14}, Y_{15})^T$$

$$\Lambda = \begin{vmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & \lambda_1 & \lambda_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \lambda_3 & \lambda_4 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \lambda_5 & \lambda_6 & \lambda_7 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{vmatrix}^T$$

$$\epsilon = (\epsilon_1, \epsilon_5, \epsilon_6, \epsilon_7, \epsilon_8, \epsilon_9, \epsilon_{10}, \epsilon_{11}, \epsilon_{12}, \epsilon_{13}, \epsilon_{14}, \epsilon_{15})^T$$

$$\eta = (\eta_1, \eta_5, \eta_6, \eta_7, \eta_8)^T$$

$$\text{with } \text{cov}(\epsilon_5, \epsilon_6) = \theta_{56} \neq 0,$$

$$\text{cov}(\epsilon_8, \epsilon_9) = \theta_{89} \neq 0, \text{ and } \lambda_3 = \lambda_4$$

All other covariances between the ϵ 's equal zero. The equations $Y_2 = \eta_2$, $Y_3 = \eta_3$ and $Y_4 = \eta_4$ can be neglected for the identification of the measurement model because these η 's are directly observable.

This measurement model is identified if the covariance matrix of the η 's ($\text{var}(\eta)$), Λ and θ_e can be determined uniquely from the covariance matrix of the y 's. The identification of $\text{VAR}(\eta)$ can then be used to identify the structural part of the whole model.

Let $(\sigma_{ij})_{i,j=1,\dots,15}$ be the correlation matrix of the observed y -variables. Then,

$\text{cov}(\eta_1, \eta_5)$, $\text{cov}(\eta_1, \eta_6)$, $\text{cov}(\eta_1, \eta_7)$, $\text{cov}(\eta_1, \eta_8)$, $\text{cov}(\eta_5, \eta_6)$, $\text{cov}(\eta_5, \eta_7)$, $\text{cov}(\eta_5, \eta_8)$, $\text{cov}(\eta_6, \eta_7)$, $\text{cov}(\eta_6, \eta_8)$ and $\text{cov}(\eta_7, \eta_8)$ are determined by $\sigma_{1,5}$, $\sigma_{1,10}$, $\sigma_{1,11}$, $\sigma_{1,15}$, $\sigma_{5,10}$, $\sigma_{5,11}$, $\sigma_{5,15}$, $\sigma_{10,15}$ and $\sigma_{11,15}$ respectively.

Next we determine

$$\lambda_5 = \frac{\sigma_{5,12}}{\text{cov}(\eta_5, \eta_7)} = \frac{\sigma_{5,12}}{\sigma_{5,11}}, \quad \lambda_6 = \frac{\sigma_{5,13}}{\text{cov}(\eta_5, \eta_7)} = \frac{\sigma_{5,13}}{\sigma_{5,11}} \text{ and } \lambda_7 = \frac{\sigma_{5,14}}{\text{cov}(\eta_5, \eta_7)} = \frac{\sigma_{5,14}}{\sigma_{5,11}}.$$

$$\text{Substituting } \lambda_5 \text{ and } \text{cov}(\eta_5, \eta_7) \text{ into } \lambda_1 = \frac{\sigma_{6,12}}{\lambda_5 \text{cov}(\eta_5, \eta_7)} \text{ and } \lambda_2 = \frac{\sigma_{7,12}}{\lambda_5 \text{cov}(\eta_5, \eta_7)}$$

$$\text{yields } \lambda_1 = \frac{\sigma_{6,12}}{\sigma_{5,12}} \text{ and } \lambda_2 = \frac{\sigma_{7,12}}{\sigma_{5,12}}. \text{ Now we can determine } \text{var}(\eta_1) = 1, \text{var}(\eta_5) = -\frac{\sigma_{5,7}}{\lambda_2} =$$

$$= -\frac{\sigma_{5,7} \sigma_{5,12}}{\sigma_{7,12}}, \text{var}(\eta_7) = \frac{\sigma_{11,12}}{\lambda_5} = \frac{\sigma_{11,12} \sigma_{5,11}}{\sigma_{5,12}} \text{ and } \text{var}(\eta_8) = 1.$$

From $\sigma_{1,8} = -\lambda_3 \text{cov}(\eta_1, \eta_6)$, $\sigma_{9,10} = -\lambda_4 \text{var}(\eta_6)$ and $\sigma_{8,10} = -\lambda_3 \text{var}(\eta_6)$, it follows that λ_3 , λ_4 and $\text{var}(\eta_6)$ are determined. Hence Λ and $\text{var}(\eta)$ are identified as well as θ_e , the variance-covariance matrix of the ϵ 's. Hence the measurement model is identified.

3. The second part is a structural simultaneous (sub)model specifying the relations between η_5 and η_7

$$\eta_5 = Y_1^T X + Y_5^T X_7 + Y_5^T X_8 + \beta_5^T \eta_7 + \xi_5$$

$$\eta_7 = Y_2^T X + Y_7^T X_2 + \beta_7^T \eta_5 + \xi_7$$

$$X = (X_1, X_3, X_4, X_5, X_6)^T$$

$$\text{where } \begin{vmatrix} Y_1^T \\ Y_2^T \end{vmatrix} = \begin{vmatrix} Y_{51} & Y_{53} & Y_{54} & Y_{55} & Y_{56} \\ Y_{71} & Y_{73} & Y_{74} & Y_{75} & Y_{76} \end{vmatrix}$$

Let C be the matrix

$$\begin{vmatrix} 1 & -B_{57} & Y_1^T & 0 & Y_{57} & Y_{58} \\ B_{71} & 1 & Y_2^T & Y_{72} & 0 & 0 \end{vmatrix}$$

According to the rank condition, identification is established if the matrices C_1 and C_2 consist of columns of C for which the first respectively the second row are zero elements. In this case

$$C_1 = \begin{vmatrix} 0 \\ Y_{72} \end{vmatrix} \text{ and } C_2 = \begin{vmatrix} Y_{57} & Y_{58} \\ 0 & 0 \end{vmatrix}$$

Hence, if $Y_{72} = 0$ and $Y_{57} = 0$ or $Y_{58} = 0$, then the model is identified.

4. The third part consists of the remaining structural equations. This part can be rewritten in the form $Ay = \Gamma x + \xi$, with A an upper triangular matrix with ones as diagonal elements and $\text{var}(\xi)$ a non-singular diagonal matrix. It is well known that such recursive models are always identified.

Because each of the three parts is identified and the covariances between the error terms of the subsystems are uncorrelated, it follows that the whole model is identified (see e.g. Wiley, 1973).

APPENDIX 6.A. Translation of the 21-item Questionnaire for Research into Experienced Health Status (GREH)

1. Is your appetite less good than you consider normal?
2. Does your stomach often feel full and bloated?
3. Do you get short of breath easily?
4. Do you often have pains in the chest or heart region?
5. Do you regularly have pains in the stomach region?
6. Do you often have an unpleasant or sweetish taste in your mouth?
7. Do you regularly suffer from palpitations?
8. Do you often feel tight in the chest?
9. Do your bones or muscles ever ache?
10. Do you often feel tired?
11. Do you often have headaches?
12. Do you often have indigestion?
13. Are you often troubled by back-ache?
14. Is your stomach regularly upset?
15. Do your arms and legs often go dead or get pins-and-needles?
16. Do you get tired sooner than you would consider normal?
17. Do you often feel dizzy?
18. Do you often feel listless?
19. Do you often have stomach trouble?
20. Do you often feel sleepy or sluggish?
21. Do you generally get up feeling tired and unrested in the morning?

CURRICULUM VITAE

Eddy, K.A. Van Doorslaer werd op 24 februari 1958 geboren te Willebroek (B). Hij behaalde het diploma Moderne Humaniora, Wetenschappelijke A aan het H. Pius X Instituut te Antwerpen in 1976. In 1980 werd hij licentiaat in de Toegepaste Economische Wetenschappen, richting bedrijfseconometrie aan de Universitaire Faculteiten St. Ignatius te Antwerpen (UFSIA). Van 1981 tot 1983 genoot hij een CIM-Fellowship van het Interuniversitair College voor Doctorale Studies in Managementwetenschappen te Brussel. Hij behaalde de M.Sc. Degree in Health Economics aan de Universiteit van York (GB) in 1982. Sedert 1983 is hij als universitair docent verbonden aan de Vakgroep Economie van de Gezondheidszorg van de Rijksuniversiteit Limburg en verrichtte daar voornamelijk onderzoek naar het gebruik van gezondheidszorgvoorzieningen.